

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 18:56:02 ; Search time 40.0461 Seconds
(without alignments)
1720.199 Million cell updates/sec

Title: US-09-985-689A-1

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAYNPVGVGPTSLAIVN 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```

1: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2247	100.0	434	23	AAW50080
2	2247	100.0	640	20	AA117090
3	2242	99.8	640	20	AA117091
4	2191	97.5	434	23	AAW50081
5	2183	97.2	639	20	AA117089
6	2155	95.9	639	20	AA117087
7	2155	95.9	640	20	AA117088
8	2143	95.4	434	23	AAW50085
9	2125.5	94.6	433	23	AAW50086

10	2125.5	94.6	641	20	AAW89547
11	2071	92.2	434	23	AAW50090
12	1998.5	88.9	433	23	AAW50084
13	1994.5	88.8	433	23	AAW50082
14	1989.5	88.5	636	20	AAW89548
15	1987.5	88.5	433	23	AAW50083
16	1986.5	88.4	433	23	AAW50083
17	1986.5	88.4	433	19	AAW61495
18	1986.5	88.4	433	20	AAW95698
19	1986.5	88.4	433	21	AAW69207
20	1986.5	88.4	433	21	AAW44619
21	1581.5	70.4	345	19	AAW62230
22	1581.5	70.4	345	18	AAW21654
23	452.5	20.1	659	18	AAW24121
24	452.5	20.1	659	20	AAW94840
25	414	18.4	412	20	AAW94836
26	414	18.4	522	18	AAW24122
27	414	18.4	522	20	AAW94838
28	414	18.4	654	20	AAW24129
29	414	18.4	659	18	AAW94841
30	401	17.8	659	18	AAW24123
31	372	16.6	545	22	ABO9483
32	366	16.3	1079	22	ABO81180
33	366	16.3	1079	24	ABO07391
34	346	15.4	520	18	AAW13666
35	346	15.4	734	18	AAW13667
36	346	15.4	823	18	AAW13668
37	338	15.0	1237	24	ABU11343
38	310.5	13.8	806	13	AAW27481
39	306.5	13.6	519	24	ABP76735
40	306.5	13.6	1938	24	ABP76678
41	304.5	13.6	903	17	AAW87007
42	304.5	13.6	1398	17	AAW87008
43	304.5	13.6	1398	18	AAW24124
44	304.5	13.6	1398	20	AAW94839
45	301.5	13.4	699	20	AAW08471

ALIGNMENTS

RESULT 1

AAW50080
ID AAW50080 standard; protein; 434 AA.

AC AAW50080;

DT 12-AUG-2002 (first entry)

DE Bacillus sp KSM-KP43 alkaline protease protein fragment.

KW Alkaline protease; detergent; laundry; bleaching; dishwasher.

OS Bacillus sp.

PN EPI209233-A2.

PD 29-MAY-2002.

PF 22-NOV-2001; 2001EP-0127851.

PR 22-NOV-2000; 2000JP-0355166.

PR 12-APR-2001; 2001JP-0114048.

XX (KAOS) KAO CORP.

XX Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;

XX Okuda M, Sasaki K;

XX WPI; 2002-437518/47.

XX New modified alkaline proteases useful in detergent compositions -

XX

XX

XX

XX

XX

XX

XX

XX

XX

Bacillus JP170 pro
Bacillus sp KSM-KP
Bacillus sp SD-521
Bacillus sp D6-(FE
Bacillus sp. alkal
Bacillus sp Y-(FER
Alkali-protease Ya
Modified Bacillus
Bacillus sp. Lion
Amino acid sequenc
Bacillus Lion Y en
Subtilase JP170 fr
Subtilase JP170.
Thermococcus prote
WO9856926 Seq ID 1
Hyperthermostable
Pyrococcus furiosu
Hyperthermostable
Pyrococcus furiosu
Hyperthermostable
Protease. Synthet
T. yonsei subcili
Transglutaminase r
Foreign protein #4
Fragment of dhpA g
Streptomyces virid
DhpA-mel chimeric
Protein encoded by
RP-III residual pr
Streptomyces virid
Streptomyces virid
Hyperthermostable
Protease. Pyrococ
Pyrococcus furiosu
WO9856926 Seq ID 6
F. balustium CP70

PS Claim 1; Page 10-11; 25pp; English.

XX This invention describes novel Bacillus sp. alkaline proteases useful in
CC detergent compositions, especially in laundry, bleaching or automatic
CC dishwasher detergents. The novel proteases have an increased detergency %
CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
CC sequence represents a fragment of the alkaline protease KP43 from
CC Bacillus sp strain KSM-KP43 which is used to create the modified protease
CC represented in AAM50090.

XX Sequence 434 AA;

Query Match 100.0%; Score 2247; DB 23; Length 434;
Best Local Similarity 100.0%; Pred. No. 3.3e-153;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSMHEAFRGKITALYALGRTN 60
DB 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSMHEAFRGKITALYALGRTN 60
QY 61 NANDTNGHGHVAGSVLGNSTNKGMAPQANLVFQSDGGLGGLPSNLQTLFSQAYS 120
DB 61 NANDTNGHGHVAGSVLGNSTNKGMAPQANLVFQSDGGLGGLPSNLQTLFSQAYS 120
QY 121 AGARIHTNSGAAVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
DB 121 AGARIHTNSGAAVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMPGTFFILSARSSLPDSSF 240
DB 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMPGTFFILSARSSLPDSSF 240
QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITTPKPSLLKAALIAGAADIGLY 300
DB 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITTPKPSLLKAALIAGAADIGLY 300
QY 301 PNGNOGWRVTLDKSLNVAYVNESSLSLSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
DB 301 PNGNOGWRVTLDKSLNVAYVNESSLSLSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
QY 361 SVTLVNDLDLVIITAPNGTQYVGNDFTPSYNDWGRNNVNFVINAPOSQGYTIEVOAYN 420
DB 361 SVTLVNDLDLVIITAPNGTQYVGNDFTPSYNDWGRNNVNFVINAPOSQGYTIEVOAYN 420
QY 421 VPVGPQTFSLAIVN 434
DB 421 VPVGPQTFSLAIVN 434

RESULT 2
AAY17090
ID AAY17090 standard; Protein; 640 AA.

XX AAY17090;
XX 20-MAR-2003 (updated)
DT 21-JUL-1999 (first entry)
XX Bacillus alkaline protease.

XX Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
KW washing composition; oxidising agent.

OS Bacillus sp.
XX WO9918218-A1.
XX 15-APR-1999.
XX 07-OCT-1998; 98WO-JP04528.
XX 07-OCT-1997; 97JP-0274570.

1020

PA (KAOS) KAO CORP.

XX Hitomi J, Kageyama Y, Kubota H, Nomura M, Okuda M;
PI Saeki K, Shikata S, Takaiwa M;
XX WPI; 1999-287736/27.
DR N-PSDB; AAX37278.

XX Alkali protease from Bacillus used in washing powders

XX Disclosure; Page 58-63; 71pp; Japanese.

XX The invention relates to alkaline proteases produced by strains of
CC Bacillus. The proteases ability to digest casein is not inhibited by
CC oleic acid and they have a high stability to oxidising agents. The
CC alkaline protease of the invention has the following properties: (a) it
CC is active over the pH range 4-13 and has at least 80% of its optimum
CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
CC its ability to digest casein is not inhibited by oleic acid; (e) it has
CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can
CC used as enzymes in washing compositions for use in automatic dishwashers
CC and for washing clothes. The stability to oxidising agents allows the
CC enzyme to be an effective component of washing compositions including
CC bleaches. The present sequence represents an alkaline protease.
CC (Updated on 20-MAR-2003 to correct DR field.)

XX Sequence 640 AA;

Query Match 100.0%; Score 2247; DB 20; Length 640;
Best Local Similarity 100.0%; Pred. No. 5.5e-153;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSMHEAFRGKITALYALGRTN 60
DB 207 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSMHEAFRGKITALYALGRTN 265
QY 61 NANDTNGHGHVAGSVLGNSTNKGMAPQANLVFQSDGGLGGLPSNLQTLFSQAYS 120
DB 267 NANDTNGHGHVAGSVLGNSTNKGMAPQANLVFQSDGGLGGLPSNLQTLFSQAYS 325
QY 121 AGARIHTNSGAAVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
DB 327 AGARIHTNSGAAVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 386
QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMPGTFFILSARSSLPDSSF 240
DB 387 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMPGTFFILSARSSLPDSSF 446
QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITTPKPSLLKAALIAGAADIGLY 300
DB 447 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITTPKPSLLKAALIAGAADIGLY 506
QY 301 PNGNOGWRVTLDKSLNVAYVNESSLSLSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
DB 507 PNGNOGWRVTLDKSLNVAYVNESSLSLSTOKATYSFTATAGKPLKISLVWSDAPASTTA 566
QY 361 SVTLVNDLDLVIITAPNGTQYVGNDFTPSYNDWGRNNVNFVINAPOSQGYTIEVOAYN 420
DB 567 SVTLVNDLDLVIITAPNGTQYVGNDFTPSYNDWGRNNVNFVINAPOSQGYTIEVOAYN 626
QY 421 VPVGPQTFSLAIVN 434
DB 627 VPVGPQTFSLAIVN 640

RESULT 3
AAY17091
ID AAY17091 standard; Protein; 640 AA.
XX AAY17091;
XX 20-MAR-2003 (updated)

Common Applicants

DT XX 21-JUL-1999 (first entry)
DE XX Bacillus alkaline protease.
KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
KW washing composition; oxidising agent.
XX XX
XX OS Bacillus sp.
XX PN WO9918218-A1.
XX PD 15-APR-1999.
XX PF 07-OCT-1998; 98WO-JP04528.
XX PR 07-OCT-1997; 97JP-0274570.
XX PA (KAOS) KAO CORP.
XX PI Hitomi J, Kageyama Y, Kubota H, Nomura M, Okuda M;
XX PI Saeiki K, Shikata S, Takaiwa M;
XX XX WPI; 1999-287736/27.
XX DR N-PSDB; AAX37279.
XX XX
XX PT Alkali protease from Bacillus used in washing powders
XX PS Disclosure; Page 63-68; 71pp; Japanese.
XX XX
CC The invention relates to alkaline proteases produced by strains of
CC Bacillus. The proteases ability to digest casein is not inhibited by
CC oleic acid and they have a high stability to oxidising agents. The
CC alkaline protease of the invention has the following properties: (a) it
CC is active over the pH range 4-13 and has at least 80% of its optimum
CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
CC stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d)
CC its ability to digest casein is not inhibited by oleic acid; (e) it has
CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
CC used as enzymes in washing compositions for use in automatic dishwashers
CC and for washing clothes. The stability to oxidising agents allows the
CC enzyme to be an effective component of washing compositions including
CC bleaches. The present sequence represents an alkaline protease.
CC (Updated on 20-MAR-2003 to correct DR field.)
XX XX
SQ Sequence 640 AA;

Query Match 99.8%; Score 2242; DB 20; Length 640;
Best Local Similarity 99.8%; Pred. No. 1.3e-152;
Matches 433; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGVIVAVADTGLDTRNDSSMHEAFRGKITALYALGRN 60
DB 207 NDVARGIVKADVAQSSYGLYGQGVIVAVADTGLDTRNDSSMHEAFRGKITALYALGRN 266

QY 61 NANDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
DB 267 NANDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 326

QY 121 AGARIHTNSWGAANVCAYTTDSRNVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
DB 327 AGARIHTNSWGAANVCAYTTDSRNVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 386

QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTILSARSLAPDSSF 240
DB 387 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTILSARSLAPDSSF 446

QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHEVKNRGITPKPSLLKAALIAGAADIGLY 300
DB 447 WANHDSKYAYMGGTSMATPIVAGNVAQLREHEVKNRGITPKPSLLKAALIAGAADIGLY 506

QY 301 PNGOGWGRVTLDKSLNVAAYNESSLSSTSQKATYSFTATAGKPLKISLWSDAPASTTA 360
DB 507 PNGOGWGRVTLDKSLNVAAYNESSLSSTSQKATYSFTATAGKPLKISLWSDAPASTTA 566

QY 361 SVTLVNDLVLVITAPNGTQYVGNDFSTPYNDNWDGRNNVNFVINAPOSGTYYTIEVOAYN 420
DB 567 SVTLVNDLVLVITAPNGTQYVGNDFSTPYNDNWDGRNNVNFVINAPOSGTYYTIEVOAYN 626

QY 421 VPVGPQTFSLAIVN 434
DB 627 VPVGPQNFSLAIVN 640

RESULT 4
AAM50081
ID AAM50081 standard; protein; 434 AA.
XX
AC AAM50081;
XX
DT 12-AUG-2002 (first entry)
XX
DE Bacillus sp KSM-KP9860 alkaline protease protein fragment.
XX
KW Alkaline protease; detergent; laundry; bleaching; dishwasher.
XX
OS Bacillus sp.
XX
PN EP1209233-A2.
XX
PD 29-MAY-2002.
XX
PF 22-NOV-2001; 2001EP-0127851.
XX
PR 22-NOV-2000; 2000JP-0355166.
XX
PR 12-APR-2001; 2001JP-0114048.
XX
PA (KAOS) KAO CORP.
XX
PI Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
PI Okuda M, Saeki K;
XX
DR WPI; 2002-437518/47.
XX
PT New modified alkaline proteases useful in detergent compositions -
XX
PS Claim 5; Page 12-13; 25pp; English.
XX
CC This invention describes novel Bacillus sp. alkaline proteases useful in
CC detergent compositions, especially in laundry, bleaching or automatic
CC dishwasher detergents. The novel proteases have an increased detergency %
CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
CC sequence represents a fragment of the alkaline protease KP9860 from
CC Bacillus sp strain KSM-KP9860 described in the method of the invention.
XX
SQ Sequence 434 AA;

Query Match 97.5%; Score 2191; DB 23; Length 434;
Best Local Similarity 96.5%; Pred. No. 3.5e-149;
Matches 419; Conservative 13; Mismatches 2; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGVIVAVADTGLDTRNDSSMHEAFRGKITALYALGRN 60
DB 1 NDVARGIVKADVAQSSYGLYGQGVIVAVADTGLDTRNDSSMHEAFRGKITALYALGRN 60

QY 61 NANDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
DB 61 NANDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120

QY 121 AGARIHTNSWGAANVCAYTTDSRNVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
DB 121 AGARIHTNSWGAANVCAYTTDSRNVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180

QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTILSARSLAPDSSF 240
DB 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTILSARSLAPDSSF 240

QY	241	WANHDSKYAYMGTSMTPIVAGNAQLRHFVKNRGITPKPSLLKALIAAGADIGLY	300
Db	241	WANHDSKYAYMGTSMTPIVAGNAQLRHFVKNRGITPKPSLLKALIAAGADVGLGY	300
QY	301	PNGNOGWRVTLDKSLNVAYVNESSLSSTOKATYFTATAGKPLKISLVNSDAPASTTA	360
Db	301	PNGNOGWRVTLDKSLNVAYVNESSLSSTOKATYFTATAGKPLKISLVNSDAPASTA	360
QY	361	SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGRNNVENVFINAPOSQYTTIEVOAYN	420
Db	361	SVTLVNDLVLITAPNGTRYVGNDFSAPFNDNDGRNNVENVFINSPOSQYTTIEVOAYN	420
QY	421	VPVGPOTFSLAIYN 434	
Db	421	VPVGPQNFSLAIYN 434	
RESULT 5			
AA17089	standard; Protein: 639 AA.		
ID	AA17089		
AC	AA17089;		
DT	20-MAR-2003 (updated)		
DT	21-JUL-1999 (first entry)		
XX	Bacillus alkaline protease.		
DE	Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;		
KW	washing composition; oxidising agent.		
XX	Bacillus sp.		
XX	WO9918218-A1.		
PN	15-APR-1999.		
PD	07-OCT-1998; 98WO-JP04528.		
PF	07-OCT-1997; 97JP-0274570.		
PR	(KAOS) KAO CORP.		
PI	Hitomi J, Kageyama Y, Kubota H, Nomura M, Okuda M;		
PI	Saeki K, Shikata S, Takaiwa M;		
XX	WPI; 1999-287736/27.		
DR	N-PSDB; AAX37277.		
XX	Alkali protease from Bacillus used in washing powders		
PT	Disclosure; Page 53-58; 71pp; Japanese.		
XX	The invention relates to alkaline proteases produced by strains of		
CC	Bacillus. The proteases ability to digest casein is not inhibited by		
CC	oleic acid and they have a high stability to oxidising agents. The		
CC	alkaline protease of the invention has the following properties: (a) it		
CC	is active over the pH range 4-13 and has at least 80% of its optimum		
CC	activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is		
CC	stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)		
CC	its ability to digest casein is not inhibited by oleic acid; (e) it has		
CC	molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be		
CC	used as enzymes in washing compositions for use in automatic dishwashers		
CC	and for washing clothes. The stability to oxidising agents allows the		
CC	enzyme to be an effective component of washing compositions including		
CC	bleaches. The present sequence represents an alkaline protease.		
CC	(Updated on 20-MAR-2003 to correct DR field.)		
XX	Sequence 639 AA;		
SQ	Query Match 97.2%; Score 2183; DB 20; Length 639;		
	Best Local Similarity 96.3%; Pred. No. 2.2e-148;		
	Matches 418; Conservative 13; Mismatches 3; Indels 0; Gaps 0;		

QY	1	NDVARGIVKADVAOSSYGLYGOGIIVAVADTGLDTRNDSSMHEAFRGRITAYALGRTN	60
Db	206	NDVARGIVKADVAOSSYGLYGOGIIVAVADTGLDTRNDSSMHEAFRGRITAYALGRTN	265
QY	61	NANDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFOSIMDSGGGLGGLPSNLQTLFSQAYS	120
Db	266	NANDTNGHGTTHVAGSVLNGATNKGMAPOANLVFOSIMDSGGGLGGLPSNLQTLFSQAFS	325
QY	121	AGARIHTNSWGAAYNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNCGTISAPCTAKNAI	180
Db	326	AGARIHTNSWGAAYNGAYTTDSRNVDYVRKNDMTILFAAGNERPNCGTISAPCTAKNAI	385
QY	181	TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTFILSARSSLAPDSSF	240
Db	386	TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTVILSARSSLAPDSSF	445
QY	241	WANHDSKYAYMGTSMTPIVAGNVAQLRHFVKNRGITPKPSLLKALIAAGADIGLY	300
Db	446	WANHDSKYAYMGTSMTPIVAGNVAQLRHFVKNRGITPKPSLLKALIAAGADVGLGY	505
QY	301	PNGNOGWRVTLDKSLNVAYVNESSLSSTOKATYFTATAGKPLKISLVNSDAPASTTA	360
Db	506	PNGNOGWRVTLDKSLNVAYVNESSLSSTOKATYFTATAGKPLKISLVNSDAPASTTA	565
QY	361	SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGRNNVENVFINAPOSQYTTIEVOAYN	420
Db	566	SVTLVNDLVLITAPNGTRYVGNDFSAPFNDNDGRNNVENVFINSPOSQYTTIEVOAYN	625
QY	421	VPVGPOTFSLAIYN 434	
Db	626	VPVGPQNFSLAIYN 639	
RESULT 6			
AA17087	standard; protein; 639 AA.		
ID	AA17087		
AC	AA17087;		
DT	20-MAR-2003 (updated)		
DT	21-JUL-1999 (first entry)		
XX	An alkaline protease sequence from Bacillus species.		
DE	Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;		
KW	washing composition; oxidising agent.		
XX	Bacillus sp.		
XX	Key Location/Qualifiers		
FT	Misc-difference 1..639 /note= "all residues indicated as Xaa are arbitrary amino acids"		
FT	WO9918218-A1.		
XX	15-APR-1999.		
XX	07-OCT-1998; 98WO-JP04528.		
XX	07-OCT-1997; 97JP-0274570.		
XX	(KAOS) KAO CORP.		
PA	Hitomi J, Kageyama Y, Kubota H, Nomura M, Okuda M;		
PI	Saeki K, Shikata S, Takaiwa M;		
XX	WPI; 1999-287736/27.		
DR	N-PSDB; AAX37277.		
XX	Alkali protease from Bacillus used in washing powders		

PS Claim 3; Page 47-50; 71pp; Japanese.
 XX The invention relates to alkaline proteases produced by strains of
 CC Bacillus. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease of the
 CC invention.
 CC (Updated on 20-MAR-2003 to correct DR field.)
 XX Sequence 639 AA;
 SQ Query Match 95.9%; Score 2155; DB 20; Length 639;
 Best Local Similarity 96.3%; Pred. No. 2.2e-146;
 Matches 418; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
 QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSMHEAFRGKITALYALGRTN 60
 DB 206 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSMHEAFRGKITALYALGRTN 265
 QY 61 NANDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGSLPSNLOTLSQAYS 120
 DB 266 NANDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGSLPSNLOTLSQAYS 325
 QY 121 AGARIHNSWGAAYNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 DB 326 AGARIHNSWGAAYNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 385
 QY 181 TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIKPDVMAPTFILLSARSLAPDSF 240
 DB 386 TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIKPDVMAPTFILLSARSLAPDSF 445
 QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKAALAGAADIGLY 300
 DB 446 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKAALAGAADIGLY 505
 QY 301 PNGNOGWRVTLDKSLNVAYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
 DB 506 PNGNOGWRVTLDKSLNVAYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 565
 QY 361 SVTLVNDLVLITAPNGTQYVGNDFTPYNDNWDGRNNVENVFINAPQSGTYTIEVQAYN 420
 DB 566 SVTLVNDLVLITAPNGTQYVGNDFTPYNDNWDGRNNVENVFINAPQSGTYTIEVQAYN 625
 QY 421 VPVGQPTFSLAIVN 434
 DB 626 VPVGQPTFSLAIVN 639
 RESULT 7
 AAY17088
 ID AAY17088 standard; protein; 640 AA.
 XX AC AAY17088;
 XX DT 20-MAR-2003 (updated)
 XX DT 21-JUL-1999 (first entry)
 DE An alkaline protease sequence from Bacillus species.
 KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent.
 XX OS Bacillus sp.
 XX

FH Key Location/Qualifiers
 FT Misc-difference 1..640
 FT /note= "all residues indicated as Xaa are arbitrary
 XX amino acids"
 PN WO9918218-AL.
 XX 15-APR-1999.
 XX 07-OCT-1998; 98WO-JP04528.
 XX 07-OCT-1997; 97JP-0274570.
 XX (KAOS) KAO CORP.
 PI Hitomi J, Kageyama Y, Kubota H, Nomura M, Okuda M;
 PI Saeki K, Shikata S, Takaiwa M;
 XX WPI: 1999-287736/27.
 XX N-PSDB; AAX37278.
 DR Alkali protease from Bacillus used in washing powders
 XX Claim 3; Page 50-53; 71pp; Japanese.
 PS The invention relates to alkaline proteases produced by strains of
 CC Bacillus. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease of the
 CC invention.
 CC (Updated on 20-MAR-2003 to correct DR field.)
 XX Sequence 640 AA;
 QY Query Match 95.9%; Score 2155; DB 20; Length 640;
 Best Local Similarity 96.3%; Pred. No. 2.2e-146;
 Matches 418; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
 QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSMHEAFRGKITALYALGRTN 60
 DB 207 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSMHEAFRGKITALYALGRTN 266
 QY 61 NANDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGSLPSNLOTLSQAYS 120
 DB 267 NANDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGSLPSNLOTLSQAYS 326
 QY 121 AGARIHNSWGAAYNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 DB 327 AGARIHNSWGAAYNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 386
 QY 181 TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIKPDVMAPTFILLSARSLAPDSF 240
 DB 387 TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIKPDVMAPTFILLSARSLAPDSF 446
 QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKAALAGAADIGLY 300
 DB 447 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKAALAGAADIGLY 506
 QY 301 PNGNOGWRVTLDKSLNVAYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
 DB 507 PNGNOGWRVTLDKSLNVAYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 566
 QY 361 SVTLVNDLVLITAPNGTQYVGNDFTPYNDNWDGRNNVENVFINAPQSGTYTIEVQAYN 420
 XX

567 SVTLVNDLVLITAPNGCTXYVGNDFXXPXXXNWDGRNNVNFVFNXPSGTYYIEVOAYN 626

301 PNGQGRVTLDKSLNVAYVNESSLSQKATYSFTATAGKPLKISLVSDAPASTTA 360
 301 PSNGQGRVTLDKSLNVAFVNETSLSNQRKATYSFTAQSGKPLKISLVSDAPASTSA 360
 361 SVTLVNDLVLITAPNGCTXYVGNDFXXPXXXNWDGRNNVNFVFNXPSGTYYIEVOAYN 420
 361 SVTLVNDLVLITAPNGCTXYVGNDFXXPXXXNWDGRNNVNFVFNXPSGTYYIEVOAYN 420

RESULT 8

AAM50085
 ID AAM50085 standard; protein; 434 AA.
 XX
 AC AAM50085;
 XX
 DT 12-AUG-2002 (first entry)
 XX
 DE Bacillus sp alkaline protease protein A-1 fragment.
 XX
 DE Alkaline protease; detergent; laundry; bleaching; dishwasher.
 KW
 KW Bacillus sp.
 OS
 XX
 XX EPI209233-A2.
 PN
 XX
 PD 29-MAY-2002.
 XX
 XX 22-NOV-2001; 2001EP-0127851.
 XX
 XX 22-NOV-2000; 2000JP-0355166.
 PR
 PR 12-APR-2001; 2001JP-0114048.
 XX
 XX (KAOS) KAO CORP.
 PA
 XX
 XX Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
 PI
 PI Okuda M, Saeki K;
 XX
 DR WPI; 2002-437518/47.
 XX
 PT New modified alkaline proteases useful in detergent compositions -
 PS
 PS Claim 5; Page 18-19; 25pp; English.
 XX
 XX This invention describes novel Bacillus sp. alkaline proteases useful in
 CC detergent compositions, especially in laundry, bleaching or automatic
 CC dishwasher detergents. The novel proteases have an increased detergency &
 CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
 CC sequence represents a fragment of the alkaline protease A-1 from
 CC Bacillus sp NCIB12289 described in the method of the invention.
 XX
 XX
 SQ Sequence 434 AA;

Query Match 95.4%; Score 2143; DB 23; Length 434;
 Best Local Similarity 93.5%; Pred. No. 9,7e-146;
 Matches 406; Conservative 19; Mismatches 9; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITIYALGRTN 60
 DB 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITIYALGRTN 60
 QY 61 NANDTNGHTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLPSNLQTLFISOAYS 120
 DB 61 NANDPNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLPSNLQTLFISOAYS 120
 QY 121 AGARIHTNSWGAAPVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
 DB 121 AGARIHTNSWGAAPVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
 QY 181 TVGATENLRPSFGSYADININHAQFSSRGPTKDGRIKPDVMAPTFILLSARSLAPDSF 240
 DB 181 TVGATENLRPSFGSYADININHAQFSSRGPTKDGRIKPDVMAPTFILLSARSLAPDSF 240
 QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITTPKPSLLKAALIAAGADIGLY 300
 DB 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITTPKPSLLKAALIAAGADIGLY 300

QY 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITIYALGRTN 60
 DB 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITIYALGRTN 60
 QY 61 NANDTNGHTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLPSNLQTLFISOAYS 120
 DB 61 NANDPNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLPSNLQTLFISOAYS 120
 QY 121 AGARIHTNSWGAAPVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
 DB 121 AGARIHTNSWGAAPVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
 QY 181 TVGATENLRPSFGSYADININHAQFSSRGPTKDGRIKPDVMAPTFILLSARSLAPDSF 240
 DB 181 TVGATENLRPSFGSYADININHAQFSSRGPTKDGRIKPDVMAPTFILLSARSLAPDSF 240
 QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITTPKPSLLKAALIAAGADIGLY 300
 DB 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITTPKPSLLKAALIAAGADIGLY 300

Query Match 94.6%; Score 2125.5; DB 23; Length 433;
 Best Local Similarity 93.5%; Pred. No. 1.7e-144;
 Matches 406; Conservative 19; Mismatches 8; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITIYALGRTN 60
 DB 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITIYALGRTN 60
 QY 61 NANDTNGHTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLPSNLQTLFISOAYS 120
 DB 61 NANDPNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLPSNLQTLFISOAYS 119
 QY 121 AGARIHTNSWGAAPVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
 DB 120 AGARIHTNSWGAAPVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 179

Query Match 94.6%; Score 2125.5; DB 23; Length 433;
 Best Local Similarity 93.5%; Pred. No. 1.7e-144;
 Matches 406; Conservative 19; Mismatches 8; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITIYALGRTN 60
 DB 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITIYALGRTN 60
 QY 61 NANDTNGHTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLPSNLQTLFISOAYS 120
 DB 61 NANDPNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLPSNLQTLFISOAYS 119
 QY 121 AGARIHTNSWGAAPVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
 DB 120 AGARIHTNSWGAAPVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 179

Query Match 94.6%; Score 2125.5; DB 23; Length 433;
 Best Local Similarity 93.5%; Pred. No. 1.7e-144;
 Matches 406; Conservative 19; Mismatches 8; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITIYALGRTN 60
 DB 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITIYALGRTN 60
 QY 61 NANDTNGHTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLPSNLQTLFISOAYS 120
 DB 61 NANDPNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLPSNLQTLFISOAYS 119
 QY 121 AGARIHTNSWGAAPVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
 DB 120 AGARIHTNSWGAAPVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 179

Query Match 94.6%; Score 2125.5; DB 23; Length 433;
 Best Local Similarity 93.5%; Pred. No. 1.7e-144;
 Matches 406; Conservative 19; Mismatches 8; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITIYALGRTN 60
 DB 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITIYALGRTN 60
 QY 61 NANDTNGHTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLPSNLQTLFISOAYS 120
 DB 61 NANDPNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLPSNLQTLFISOAYS 119
 QY 121 AGARIHTNSWGAAPVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
 DB 120 AGARIHTNSWGAAPVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 179

QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAQCTFILLSRSSLPDSSF 240
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 180 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAQCTFILLSRSSLPDSSF 239
 QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKALAGAADIGLY 300
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 240 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKALAGAADVGLGF 299
 QY 301 PNGNOGWGRVTLDKSLNVAVYNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 300 PNGNOGWGRVTLDKSLNVAVFNETSPLSTOKATYSFTAQAGKPLKISLVWSDAPGSTTA 359
 QY 361 SVTLVNDLVLITAPNGTQYVGNDEFTSPYNDNWDGRNNVENFINAPQSGTYTTEVQAYN 420
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 360 SFTLVNDLVLITAPNGTKYVGNDEFTAPYDNNWDGRNNVENFINAPQSGTYTTEVQAYN 419
 QY 421 VPVGPOTFSLAIVN 434
 Db |||||||||||||||
 420 VPVSPQTFSLAIVH 433

RESULT 10
 AAM89547
 ID AAM89547 standard; Protein; 641 AA.

XX AC AAM89547;
 XX DT 12-APR-1999 (first entry)
 XX DE Bacillus Jp170 protease.
 XX KW Protease; detergent; surfactant; leather processing; debittering;
 XX OS flavour.
 XX XX Bacillus sp.

Key Location/Qualifiers
 FT Peptide 1..33
 FT /note= "signal peptide"
 FT Region 34...208
 FT /note= "prepro region"
 FT Protein 209..641
 FT /note= "mature protein"

PN WO9856927-A2.
 XX 17-DEC-1998.
 XX PF 09-JUN-1998; 98WO-US12005.
 XX PR 12-JUN-1997; 97US-0873479.
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 PI Christianson L, Sloma A;
 XX WPI; 1999-080908/07.
 DR N-PSDB; AAV82382.

PT Novel protease from *Bacillus subtilis* LC20 - useful in laundry and
 XX dishwashing detergents and for leather processing
 PS Claim 7; Page 53-54; 77pp; English.

CC This is the amino acid sequence of a novel protease of *Bacillus* sp.
 CC Jp170 (NCIB 12513), as deduced from the nucleotide sequence of an
 CC isolated gene (see AAV82382). The entire protein, including the
 CC signal peptide and prepro region, has 77% identity to alkaline
 CC protease Y (see AAM89548) from *Bacillus*. The invention provides
 CC vectors, recombinant host cells and methods for the recombinant
 CC production of the protease. The protease is used in laundry and
 CC dishwashing detergents, for institutional and industrial cleaning,
 CC and for leather processing, as well as for debittering and

CC enhancing the degree of hydrolysis of protein hydrolysates, for
 CC flavour development through hydrolysis of proteins, degradation of
 CC undesired peptides and in enzymatic synthesis of peptides. It has
 CC enhanced stability towards oxidation under alkaline conditions,
 CC e.g. towards bleaching agents of the peroxy type. The invention
 CC also provides mutant cells in which the protease activity is
 CC diminished. Such cells can be used for the production of
 CC heterologous recombinant proteins.

XX Sequence 641 AA;

Query Match 94.6%; Score 2125.5; Db 20; Length 641;
 Best Local Similarity 93.5%; Pred. No. 2.9e-144;
 Matches 406; Conservative 19; Mismatches 8; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 209 NDVARGIVKADVAQNNFGLYGQIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 268
 QY 61 NANDTNGHGHVAGSVLNGSTNKGMAQANLVFQSIMDSGGGLGGLPSNLOTFLFSQAYS 120
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 269 NANDPNGHGHVAGSVLGN-ATNKGMAQANLVFQSIMDSGGGLGGLPANLQTLFSQAYS 327
 QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 328 AGARIHTNSGAPVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPGSGTISAPGTAKNAI 387
 QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAQCTFILLSRSSLPDSSF 240
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 388 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAQCTFILLSRSSLPDSSF 447
 QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKALAGAADIGLY 300
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 448 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKALAGAADVGLGF 507
 QY 301 PNGNOGWGRVTLDKSLNVAVYNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 508 PNGNOGWGRVTLDKSLNVAVFNETSPLSTOKATYSFTAQAGKPLKISLVWSDAPGSTTA 567
 QY 361 SVTLVNDLVLITAPNGTQYVGNDEFTSPYNDNWDGRNNVENFINAPQSGTYTTEVQAYN 420
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 568 SFTLVNDLVLITAPNGTKYVGNDEFTAPYDNNWDGRNNVENFINAPQSGTYTTEVQAYN 627
 QY 421 VPVGPOTFSLAIVN 434
 Db |||||||||||||||
 628 VPVSPQTFSLAIVH 641

RESULT 11
 AAM50090

ID AAM50090 standard; protein; 434 AA.

XX AC AAM50090;

XX DT 12-AUG-2002 (first entry)

XX DE *Bacillus* sp KSM-Kp43 alkaline protease protein variant.

XX KW Alkaline protease; detergent; laundry; bleaching; dishwasher; mutant;
 XX OS mutin.

XX OS *Bacillus* sp.
 XX OS Synthetic.

XX Key Location/Qualifiers
 XX FT Misc-difference 46

FT /label= Y,w,a,d,e,t,v,i,i,h,s,k,q,m,c
 FT /note= "as claimed in Claim 3"

FT Misc-difference 54

FT /label= Y,f,a,n,e,t,v,h,s,k,e,m,g,d,p,r,c
 FT /note= "as claimed in Claim 3"

FT Misc-difference 57
 FT /label= k,s,e,f,v,r,y,l,i,t,m,c,y,d,e,h,p,a

FT FT /note= "as claimed in Claim 3"
 FT FT 66
 FT FT /label= e,d,s,q,a,t,l,m,c,v,g,i
 FT FT /note= "as claimed in Claim 3"
 FT FT 84
 FT FT /label= OTHER, R
 FT FT /note= "OTHER= deleted residue. Specifically described
 in Claim 1"
 FT FT 101..106
 FT FT /label= k,s,e,f,v,r,y,l,i,t,m,c,y,d,e,h,p,a
 FT FT /note= "as claimed in Claim 3"
 FT FT 104
 FT FT /label= OTHER, P
 FT FT /note= "OTHER= deleted residue. Specifically described
 in Claim 1"
 FT FT 107
 FT FT /label= k,r,a,s
 FT FT /note= "as claimed in Claim 3"
 FT FT 119
 FT FT /label= y,f,a,n,e,t,v,h,s,k,e,m,g,d,p,r,c
 FT FT /note= "as claimed in Claim 3"
 FT FT 124
 FT FT /label= a,k
 FT FT /note= "as claimed in Claim 3"
 FT FT 136
 FT FT /label= k,s,e,f,v,r,y,l,i,t,m,c,y,d,e,h,p,a
 FT FT /note= "as claimed in Claim 3"
 FT FT 138
 FT FT /label= y,f,a,n,e,t,v,h,s,k,e,m,g,d,p,r,c
 FT FT /note= "as claimed in Claim 3"
 FT FT 148
 FT FT /label= y,f,a,n,e,t,v,h,s,k,e,m,g,d,p,r,c
 FT FT /note= "as claimed in Claim 3"
 FT FT 193
 FT FT /label= k,s,e,f,v,r,y,l,i,t,m,c,y,d,e,h,p,a
 FT FT /note= "as claimed in Claim 3"
 FT FT 195
 FT FT /label= y,f,a,h,e,t,v,h,s,k,e,m,g,d,p,r,c
 FT FT /note= "as claimed in Claim 3"
 FT FT 205
 FT FT /label= y,w,a,d,e,t,v,l,i,h,s,k,q,m,c
 FT FT /note= "as claimed in Claim 3"
 FT FT 247
 FT FT /label= w,f,a,r,e,t,v,l,i,h,s,e,m,c
 FT FT /note= "as claimed in Claim 3"
 FT FT 256
 FT FT /label= OTHER,a,s,e,v,l,r,e,d
 FT FT /note= "OTHER= deleted residue. Specifically described
 in Claim 1"
 FT FT 264
 FT FT /label= e,d,s,q,a,t,l,m,c,v,g,i
 FT FT /note= "as claimed in Claim 3"
 FT FT 257
 FT FT /label= v,i
 FT FT /note= "as claimed in Claim 3"
 FT FT 342
 FT FT /label= k,s,e,f,v,r,y,l,i,t,m,c,y,d,e,h,p,a
 FT FT /note= "as claimed in Claim 3"
 FT FT 369
 FT FT /label= OTHER, d
 FT FT /note= "OTHER= deleted residue. Specifically described
 in Claim 1"
 FT FT
 XX EP1209233-A2.
 XX PN
 XX PD
 XX 29-MAY-2002.
 XX PF 22-NOV-2001; 2001EP-0127851.
 XX PR 22-NOV-2000; 2000JP-0355166.
 XX PR 12-APR-2001; 2001JP-0114048.
 XX (KAOS) KAO CORP.

XX Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
 PI Okuda M, Saeki K;
 XX WPI; 2002-437518/47.
 XX New modified alkaline proteases useful in detergent compositions -
 PT Claim 1; Page -: 25pp; English.
 XX This invention describes novel Bacillus sp. alkaline proteases useful in
 CC detergent compositions, especially in laundry, bleaching or automatic
 CC dishwasher detergents. The novel proteases have an increased detergency
 CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
 CC sequence represents a variant of the alkaline protease Kp43 from
 CC Bacillus sp strain KSM-Kp43 created from the wild-type protease
 CC represented in AAM50080
 CC NOTE: This sequence is not represented in the specification but has
 CC been constructed from the sequence represented in record AAM50080.
 XX SQ Sequence 434 AA;
 Query Match 92.2%; Score 2071; DB 23; Length 434;
 Best Local Similarity 94.0%; Pred. No. 1.4e-140;
 Matches 408; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
 QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSMHEAFRGKITALYALGRTN 60
 DB 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSMHEAXRGKITALXALXRTN 60
 QY 61 NANDTNGHGTHTVAGSVLNGSTNKGMAPQANLVFQSIMDSGGGLGSLNQLTFSQAYS 120
 DB 61 NANDTNGHGTHTVAGSVLNGSTNKGMAPQANLVFQSIMDSGGGLGSLNQLTFSQAYS 120
 QY 121 AGARIHTNSWGAANGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 DB 121 AGAXIHTNSWGAANGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 QY 181 TVGATENLRPFGSYADNINHVAFSSRGPTKDGRIKPDVMPAGTIFILSARSLAPDSSF 240
 DB 181 TVGATENLRPFGSYADNINHVAFSSRGPTKDGRIKPDVMPAGTIFILSARSLAPDSSF 240
 QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
 DB 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
 QY 301 PNGNQGWGRVTLDKSLNVAYVNESSISTSKATYSFTATACKPLKISLVWSDAPASTTA 360
 DB 301 PNGNQGWGRVTLDKSLNVAYVNESSISTSKATYSFTATACKPLKISLVWSDAPASTTA 360
 QY 361 SVTLVNDLXLVITAPNGTOYVGNDFTSPYNDWGRNNVNFIAPQSGTITIEVOAYN 420
 DB 361 SVTLVNDLXLVITAPNGTOYVGNDFTSPYNDWGRNNVNFIAPQSGTITIEVOAYN 420
 QY 421 VPVGQPTFSLAIVN 434
 DB 421 VPVGQPTFSLAIVN 434
 RESULT 12
 AAM50084
 ID AAM50084 standard; protein; 433 AA.
 XX
 AC AAM50084;
 XX
 XX 12-AUG-2002 (first entry)
 DT Bacillus sp SD-521 (FERM BP-11162) alkaline protease protein fragment.
 DE Alkaline protease; detergent; laundry; bleaching; dishwasher.
 KW Bacillus sp.
 XX
 OS
 XX

PN EP1209233-A2.
 XX 29-MAY-2002.
 XX 22-NOV-2001; 2001EP-0127851.
 XX 22-NOV-2000; 2000JP-0355166.
 PR 12-APR-2001; 2001JP-0114048.
 XX (KAOS) KAO CORP.
 XX Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
 PI Okuda M, Saeki K;
 XX WPI; 2002-437518/47.
 XX New modified alkaline proteases useful in detergent compositions -
 PT Claim 5; Page 16-18; 25pp; English.
 XX This invention describes novel Bacillus sp. alkaline proteases useful in
 CC detergent compositions, especially in laundry, bleaching or automatic
 CC dishwasher detergents. The novel proteases have an increased detergency &
 CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
 CC sequence represents a fragment of the alkaline protease SD-521 from
 CC Bacillus sp strain SD-521 (FERM BP-11162) described in the method of the
 CC invention.
 XX Sequence 433 AA;
 SQ
 Query Match 88.9%; Score 1998.5; DB 23; Length 433;
 Best Local Similarity 87.8%; Pred. No. 2.3e-135;
 Matches 381; Conservative 28; Mismatches 24; Indels 1; Gaps 1;
 QY 1 NDVARGIVKADVAQSSYGLYGOGQIVAVADTGLDTGRNDSSMHEAFRGKITALYALGRTN 60
 DB 1 NDVARGIVKADVAQNNYGLYGOGQVAVADTGLDTGRNDSSMHEAFRGKITALYALGRTN 60
 QY 61 NANDPNGHGTTHVAGSVLGNSTKMGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
 DB 61 NANDPNGHGTTHVAGSVLGN-ALNKGMAPOANLVFQSIMDSGGGLGGLPSNLTLFSQAWN 119
 QY 121 AGARHTNSWGAAGVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
 DB 120 AGARHTNSWGAAPVNGAYTANSRQVDEYVRNNDMTVLEAAGNEGPNNGTISAPGTAKNAI 179
 QY 181 TVGATENLRPSFGSYADNINHAQFSSRGPTKDGRIKPDVMAAGTIFLSARSSLAPDSSF 240
 DB 180 TVGATENYRPSFGSLADPNHIAQFSSRGATRDGRIKPDVTAPGTIFLSARSSLAPDSSF 239
 QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLY 300
 DB 240 WANYSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGATDVLGY 299
 QY 301 PNGNOGWRVTLDKSLNVAAYNNESSLSSTOKATYSFTATAGKPKLKSILVWSDAPASTTA 360
 DB 300 PSGDQGWGRVTLDKSLNVAAYNEATALTGOKATYSFQAGKPKLKSILVWTDAPGSTTA 359
 QY 361 SVTLVNDLDELVTAPNGQYVGNDFTSYNDNWDGRNNVNFVINAPOSQGTYYIEVOAYN 420
 DB 360 SYTLVNDLDELVTAPNGQYVGNDFSYNDNWDGRNNVNFVINAPOSQGTYYIEVOAYN 419
 QY 421 VPVGPQTFLSAIVN 434
 DB 420 VPSGPQRFSLAIVH 433
 RESULT 13
 AAM50082
 ID AAM50082 standard; protein: 433 AA.
 XX
 AC
 XX AAM50082;

DT 12-AUG-2002 (first entry)
 XX Bacillus sp D6-(FERM P1592) alkaline protease protein fragment.
 DE Alkaline protease; detergent; laundry; bleaching; dishwasher.
 KW Bacillus sp.
 OS EP1209233-A2.
 PN 29-MAY-2002.
 XX 22-NOV-2001; 2001EP-0127851.
 XX 22-NOV-2000; 2000JP-0355166.
 PR 12-APR-2001; 2001JP-0114048.
 XX (KAOS) KAO CORP.
 XX Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
 PI Okuda M, Saeki K;
 XX WPI; 2002-437518/47.
 XX New modified alkaline proteases useful in detergent compositions -
 PT Claim 5; Page 13-15; 25pp; English.
 XX This invention describes novel Bacillus sp. alkaline proteases useful in
 CC detergent compositions, especially in laundry, bleaching or automatic
 CC dishwasher detergents. The novel proteases have an increased detergency &
 CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
 CC sequence represents a fragment of the alkaline protease E-1 from
 CC Bacillus sp strain D6-(FERM-P1592) described in the method of the
 CC invention.
 XX Sequence 433 AA;
 SQ
 Query Match 88.8%; Score 1994.5; DB 23; Length 433;
 Best Local Similarity 87.6%; Pred. No. 4.5e-135;
 Matches 380; Conservative 28; Mismatches 25; Indels 1; Gaps 1;
 QY 1 NDVARGIVKADVAQSSYGLYGOGQIVAVADTGLDTGRNDSSMHEAFRGKITALYALGRTN 60
 DB 1 NDVARGIVKADVAQNNYGLYGOGQVAVADTGLDTGRNDSSMHEAFRGKITALYALGRTN 60
 QY 61 NANDPNGHGTTHVAGSVLGNSTKMGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
 DB 61 NANDPNGHGTTHVAGSVLGN-ALNKGMAPOANLVFQSIMDSGGGLGGLPSNLTLFSQAWN 119
 QY 121 AGARHTNSWGAAGVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
 DB 120 AGARHTNSWGAAPVNGAYTANSRQVDEYVRNNDMTVLEAAGNEGPNNGTISAPGTAKNAI 179
 QY 181 TVGATENLRPSFGSYADNINHAQFSSRGPTKDGRIKPDVMAAGTIFLSARSSLAPDSSF 240
 DB 180 TVGATENYRPSFGSIADPNHIAQFSSRGATRDGRIKPDVTAPGTIFLSARSSLAPDSSF 239
 QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLY 300
 DB 240 WANYSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGATDVLGY 299
 QY 301 PNGNOGWRVTLDKSLNVAAYNNESSLSSTOKATYSFTATAGKPKLKSILVWSDAPASTTA 360
 DB 300 PSGDQGWGRVTLDKSLNVAAYNEATALTGOKATYSFQAGKPKLKSILVWTDAPGSTTA 359
 QY 361 SVTLVNDLDELVTAPNGQYVGNDFTSYNDNWDGRNNVNFVINAPOSQGTYYIEVOAYN 420
 DB 360 SYTLVNDLDELVTAPNGQYVGNDFSYNDNWDGRNNVNFVINAPOSQGTYYIEVOAYN 419
 QY 421 VPVGPQTFLSAIVN 434
 DB 420 VPSGPQRFSLAIVH 433

Db 61 NASDPNGHGHVAGSVLGN-ALNKGMAPQANLVFQSIMDSSGGLGPLSNLTLFSQAWN 119
QY 121 AGARIHTNSWGAANGAYTTDSRNVDYVRKNDWTILFAAGNEGPNGGTISAPGTAKNAI 180
Db 120 AGARIHTNSWGAAPVNGAYTANSRQVDEYVERNNDWTVLFAAGNEGPNSGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSYADNHNHVAQFSSRGPTKDGRIKPDVMTAPGTFILSARSSLAPDSSF 240
Db 180 TVGATENYRPSFGSIADNPNHIAQFSSRGATRDGRIKPDVMTAPGTFILSARSSLAPDSSF 239
QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKRCITPKPSLLKAALIAGAADICLGY 300
Db 240 WANYNSKYAYMGTSMATPIVAGNVAQLREHFVKRCITPKPSLLKAALIAGATDVGLGY 299
QY 301 PNGNOGWGRVTLDKSLNVAYVNESSLSSTOKATYSFTATAGKPLKISLVMSDAPASTTA 360
Db 300 PNGDOGWGRVTLNKSINVAYVNEATATGQKATYSFQAQAGKPLKISLVWTDAPGSTTA 359
QY 361 SVTLVNDLVLITAPNGTOYVGNDFISPYNDNWDGRRNNVENFINAPQSGTYTIEVOAYN 420
Db 360 SVTLVNDLVLITAPNGOKYVGNDFISPYDNNWDGRRNNVENFINAPQSGTYTIEVOAYN 419
QY 421 VPGPQTFSLAIVN 434
Db 420 VPSGPQFSLAIVH 433

Search completed: July 25, 2003, 18:59:39
Job time : 43.0461 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 28, 2003, 01:10:39 ; Search time 375.933 Seconds
(without alignments)
3116.395 Million cell updates/sec

Title: US-09-985-689A-1

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAYNVDPVGPQTSLAIVN 434

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame-p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US0985689/runat_25072003_143031_821/app_query.fasta_1.1166
-DB=N.Geneseq.19Jun03 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0 -1-LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MAPFIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MTN=0 -ALIGN=15
-MODE=LOCAL -OUTPMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0985689 -ECGN_1_606 -runat_25072003_143031_821 -NCPU=3
-NO_MMAP -LARGEQUERY -NRG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N.Geneseq.19Jun03.*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2247	100.0	1923	20	AA337278	Bacillus alkaline
2	2242	99.8	1923	20	AA337279	Bacillus alkaline
3	2183	97.2	1920	20	AA337277	Bacillus alkaline
4	2125.5	94.6	3003	20	AA328382	Bacillus JPI70 pro
5	1986.5	88.4	1293	13	AAQ27516	Alkali-protease ya
6	452.5	20.1	1977	18	AA325667	Thermococcus prote
7	452.5	20.1	1977	20	AA305926	W09856926 Seq ID 1
8	416.5	18.5	1236	20	AA305920	Hyperthermostable
9	416.5	18.5	1566	18	AA305668	Pyrococcus furiosu
10	416.5	18.5	1962	18	AA305695	Pyrococcus furiosu
11	416.5	18.5	1962	20	AA305929	Hyperthermostable
12	403.5	18.0	1977	18	AA305669	Protease coding se
13	372	16.6	2121	23	ABU54900	T. yonsei subtile
14	346	15.4	2539	18	AA305920	Streptomyces virid
15	346	15.4	2809	18	AA305920	DhpA-mel chimeric
16	338	15.0	135638	25	AB334289	S. atroolivaceus l
17	332.5	14.8	1329	24	ABK74643	Bacillus lichenifo
18	310.5	13.8	2532	13	AAQ29134	Encodes RP-III res
19	307.5	13.7	1859	18	AA305677	Thermococcus prote
20	307	13.7	2835	17	AA308141	Hyperthermostable
21	307	13.7	4765	17	AA308132	Protease gene. Py
22	307	13.7	4765	18	AA305670	Pyrococcus furiosu
23	307	13.7	4765	20	AA305921	Pyrococcus furiosu
24	306.5	13.6	1560	25	AB237569	Streptomyces virid
25	306.5	13.6	59816	25	AB237515	Streptomyces virid
26	306.5	13.6	59816	25	AB237516	Streptomyces virid
27	301.5	13.4	3413	20	AA305921	F. balustinum Cp70
28	296	13.2	898	17	AA308131	Hyperthermostable
29	290.5	12.9	564	17	AA308134	DNA sequence. Pyr
30	290.5	12.9	564	18	AA305676	Thermococcus prote
31	276	12.3	1306	24	ABU55784	Bacillus sp. rti45
32	276	12.3	1330	24	ABU55787	Bacillus sp. prote
33	275	12.2	346	23	ABU55453	T. yonsei subtile
34	265.5	11.8	2273	11	AAQ04339	Aquaricin I gene.
35	265.5	11.8	2274	12	AAQ12838	Aquaricin I coding
36	265.5	11.7	2274	16	AAQ75859	Aquaricin I gene.
37	263	11.7	4740	25	AB258957	Group B Streptococ
38	256.5	11.4	10216	17	AA305929	Transposon deliver
39	256	11.4	1485	24	ABK74647	Bacillus lichenifo
40	256	11.4	4650	24	ABN71526	Streptococcus poly
41	256	11.4	4710	24	ABN71162	Streptococcus poly
42	256	11.4	4734	24	ABN70525	Streptococcus poly
43	256	11.4	4770	24	ABN69191	Streptococcus poly
44	256	11.4	2155561	24	ABN71527	Streptococcus poly
45	253.5	11.3	3117	20	AA305926	wprA gene. Bacill

ALIGNMENTS

RESULT 1
AA337278
ID AA337278 standard; DNA; 1923 BP.
XX
AC AA337278;
XX
DT 20-MAR-2003 (updated)
DT 21-JUL-1999 (first entry)
XX
DE Bacillus alkaline protease encoding DNA.
XX
KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
KW washing composition; oxidising agent; ss.
XX
OS Bacillus sp.
XX
PN W09918218-AL.)

XX 15-APR-1999.
 PD 07-OCT-1998; 98WO-JP04528.
 PF 07-OCT-1997; 97JP-0274570.
 PR (KAOS) KAO CORP.
 PA
 PI Hitomi J, Kageyama Y, Kubota H, Nomura M, Okuda M;
 PI Saei K, Shikata S, Takaiwa M;
 PI WPI; 1999-287736/27.
 DR P-PSDB; AAY17088, AAY17090.
 DR
 XX Alkali protease from Bacillus used in washing powders
 PT
 XX Disclosure; Page 58-63; 7lpp; Japanese.
 PS
 CC The invention relates to alkaline proteases produced by strains of
 CC Bacillus. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidizing agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidizing agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease encoding
 CC DNA.
 CC (Updated on 20-MAR-2003 to correct DR field.)
 CC
 XX Sequence 1923 BP; 576 A; 418 C; 475 G; 454 T; 0 other;
 SQ

Alignment Scores:
 Pred. No.: 4,2e-164 Length: 1923
 Score: 2247.00 Matches: 434
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 20 Gaps: 0

US-09-985-689A-1 (1-434) x AAX37278 (1-1923)

QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 DB 619 AATGATGTTGGCGGTGAATTTGTCAAGCGGATGTGGCTCAGGAGCAGCTACGGGTGTAT 678
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAspSer 40
 DB 679 GGACAGGACAGATCGTAGCGGTTCGCCATACAGGCTGTATACAGGTCCCAATGACAGT 738
 QY 41 SerMethisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 DB 739 TCGATGATGAGCCTTCGGCGGGAATTTACTGCATTATATGCTTGGACGACGAAT 798
 QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 DB 799 AATGCCAATGATACGAATGGTTCATGCGCATGTGGCTGGCTCCGCTATTAGGAACGGC 858
 QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 DB 859 TCCACTAATAAGGAATGGCGCTCAGGGGAATCTAGTCTTCCAATCTATCATGATAGC 918
 QY 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 DB 919 GGTGGGACCTTGGAGGACTACCTTCGAATCTGCAACCTTATTCAGCCCAAGCATACAGT 978
 QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140

DB 979 GCTGGTGCCAGAAATTCATACAAACTCCTGGGAGCAGCAGTGAATGGGCTTACACAACA 1038
 QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 DB 1039 GATTCCAGAAATGTGATGACTATGTGCGCAAAATGATATGACGATCCTTTTCGTGCC 1098
 QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 DB 1099 GGAATGAAGGACCGAAGCGGGAACCATCATGCTCAGTCAGCAGCAGCAGTAAAAATGCAATA 1158
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 DB 1159 ACAGTCGGAGCTACGGGAAACCTCCGCCAAAGCTTTGGGTCTCTATGCGGACAAATATCAAC 1218
 QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 DB 1219 CATGTGGCAGATTCTCTACGTGGCCGCAAGAGGATGGAGGATCAACCGGATGTC 1278
 QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 DB 1279 ATGCACCGGGAAGCTTCATACATCATCAGCAAGATCTCTCTTGCACCGGATTCCTCTTC 1338
 QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 DB 1339 TGGCGGAACCATGACAGTAAATATGCATACATGGGTGGAACTGCCATGGCTACCCGATC 1398
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 DB 1399 GTTGCTGGAAACGTGGCAGACGCTCGTGAGCATTTGTGAAAAACAGAGGCATCACACCA 1458
 QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 DB 1459 AAGCCTTCTCTATTAAAAAGCGCACTGATTGGCGGTGCACTGACATCGGCCCTGGCTAC 1518
 QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 DB 1519 CCGAACGGTAACCAAGGATGGGACGAGTGACATTTGGATAAATCCCTGAACGTTGCCTAT 1578
 QY 321 ValAsnGluSerSerSerSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 DB 1579 GTGAACGAGTCCAGTTCTATCCACAGCAAAAAGCGACGACTCGTTTACTGTCTACT 1638
 QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 DB 1639 GCCGGCAAGCCTTTGAAAATCTCCCTGGTATGCTGTATGCCCTCGCAGCAGCAACTGCT 1698
 QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 DB 1699 TCCGTAAAGCTGTCAATGATCTGGACCTTGTCAATACCGCTCCAAATGGCACAGTAT 1758
 QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
 DB 1759 GTAGGAATGACTTTACTTCGCATACAATGATACTGGATGGCCCGCCCAATAACCTATAAC 1818
 QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 DB 1819 AATGTATTATTATTCACCAACAAACGGGACGCTATACATTTGAGGTACAGCTTATAAC 1878
 QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 DB 1879 GTACCGGTGGACACAGACCTTCTCGTTGGCAATTTGTGAAT 1920

RESULT 2
 AAX37278
 ID AAX37279 standard; DNA; 1923 BP.
 XX
 AC AAX37279;
 XX
 DT 20-MAR-2003 (updated)
 DT 21-JUL-1999 (first entry)
 XX
 DE Bacillus alkaline protease encoding DNA.
 XX

XX	DT	20-MAR-2003 (updated)	Db	796	AATCGAATGATACGAAACGGTATGGTACCCATGTGGCAGGTTCCGATTTAGGAAATGGC	855
XX	DT	21-JUL-1999 (first entry)	QY	81	SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer	100
XX	DE	Bacillus alkaline protease encoding DNA.	Db	856	GCAACGAATAAAGGAATGGACCTCAAGCAATCTGGTTTCAATCCATCATGATAGC	915
XX	XX	Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme; washing composition; oxidising agent; ss.	QY	101	GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer	120
XX	OS	Bacillus sp.	Db	916	AGTGGTGGCTGGAGGCTTCCCAATCTGCAAACTTATTCAGCCAAAGCATTCAGT	975
XX	XX	W09918218-A1.	QY	121	AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr	140
XX	PN	15-APR-1999.	Db	976	GCAGGTGCCAGAAATTCATAACAACCTCTGGGGGCGAGCGGTGAATGGGGCTACACGACA	1035
XX	PD	07-OCT-1998; 98WO-JP04528.	QY	141	AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla	160
XX	PF	07-OCT-1997; 97JP-0274570.	Db	1036	GATTCAGAAATGTGATGACTATGTAAGAAATGATATGACGATTCCTTTTCGGCGCT	1095
XX	PR	(KAOS) KAO CORP.	QY	161	GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle	180
XX	PA	Hitomi J, Kageyama Y, Kubota H, Nomura M, Okuda M;	Db	1096	GGGAATGAAAGCCGCAACGGGTACCATCAGTGCACCTGGTACGGCTAAACGCCATA	1155
XX	PI	Saeki K, Shikata S, Takaiwa M;	QY	181	ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn	200
XX	PI	WPI: 1999-287736/27.	Db	1156	ACAGTCGGCGCAACCGAAACCTGCGTCCCAAGCTTCGGTTCCTATGCAGATAATATTAAC	1215
XX	DR	P-PSDB: AAY17087, AAY17089.	QY	201	HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal	220
XX	XX	Alkali protease from Bacillus used in washing powders	Db	1216	CACGTTGCACAGTCTCTTCCCGTGGCCGACAAAGATGGCGAATCAAGCCTGATGTC	1275
XX	XX	Disclosure; Page 53-58; 71pp; Japanese.	QY	221	MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe	240
XX	CC	The invention relates to alkaline proteases produced by strains of	Db	1276	ATGGCGCCAGGACATACATATTTATCAGCAAGATCTCTCTGTCACCGGATTCCTCTTC	1335
XX	CC	Bacillus. The proteases ability to digest casein is not inhibited by	QY	241	TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle	260
XX	CC	oleic acid and they have a high stability to oxidising agents. The	Db	1336	TGGCGGAATCATGACAGCAATATGCCTATATGGTGGAGCTCCATGGCAACCCGATT	1395
XX	CC	alkaline protease of the invention has the following properties: (a) it	QY	261	ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro	280
XX	CC	is active over the pH range 4-13 and has at least 80% of its optimum	Db	1396	GTTGGCGGAATTTGCACAGCTCCGTCAGCATTTTGTGAAAATAGAGGAATCACTCT	1455
XX	CC	activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is	QY	281	LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr	300
XX	CC	stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)	Db	1456	AAGCCTTCCCTATTGAAAGCAGCTTTGATTGCAGTGCTGCTGATGTTGGATTGGTTAT	1515
XX	CC	its ability to digest casein is not inhibited by oleic acid; (e) it has	QY	301	ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr	320
XX	CC	molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be	Db	1516	CCGACGGAACCAAGGATGGGCGGAGTGGGATAATCTGTAATCTGTAATCTGCTAT	1575
XX	CC	used as enzymes in washing compositions for use in automatic dishwashers	QY	321	ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr	340
XX	CC	and for washing clothes. The stability to oxidising agents allows the	Db	1576	GTGAACGAATCCAGTGCCTTATCACTAGCCAAAAGCGACATATACCTTTACTGCAACG	1635
XX	CC	enzyme to be an effective component of washing compositions including	QY	341	AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla	360
XX	CC	bleaches. The present sequence represents an alkaline protease encoding	Db	1636	GCGGCAAGCCATTGAAATCTCCCTGCTATGCTGCTGATGCTGCTGATGCTGCTGCT	1695
XX	CC	DNA.	QY	361	SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr	380
XX	CC	(Updated on 20-MAR-2003 to correct DR field.)	Db	1696	TCGTGTAACCTGGTCAATGATTTGGATTTGGTTCATTCAGCACCAACCAACGACATAT	1755
XX	XX	Sequence 1920 BP; 576 A; 411 C; 473 G; 460 T; 0 other;	QY	381	ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu	400
XX	XX	Alignment Scores:	Db	1756	GTCGGGAATGACTCTCAGCACCATTTGACAAATACTGGGATGGCGGCAATAACCTAGAA	1815
XX	XX	Pred. No.: 3,62e-159 Length: 1920	QY	401	AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn	420
XX	XX	Score: 2183.00 Matches: 418	Db	1816	AATCTATTTATTAATTCGCCCAAGAGTGGAAACATATACCATTCAGGTGCAAGCATATAAT	1875
XX	XX	Percent Similarity: 99.31% Conservative: 13	QY	421	ValProValGlyProGlnThrPheSerLeuAlaIleValAsn	434
XX	XX	Best Local Similarity: 96.31% Mismatches: 3	Db	1876	GTGCGGCTGGACCAACAACTCTCTGTTGGCAATTTGTGAAC	1917
XX	XX	Query Match: 97.15% Indels: 0				
XX	XX	DB: 20 Gaps:				
XX	XX	US-09-985-689A-1 (1-434) x AAX37277 (1-1920)				
XX	QY	1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr				
XX	Db	616 AATGATGTGGCCAGAGGATTTGTCAAGCGGATGTGGCAGACAGCAGCTACGGTTGTAT				
XX	QY	21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer				
XX	Db	676 GGACAAGGCCAGATTGTCGAGTGGCGGATGCTGATGCTGATGCTGATGCTGATGCT				
XX	QY	41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn				
XX	Db	736 TCGATGATGAAGCCCTCCCGGTGTAATAACACGACTATATGACTGCGGCGGAGAT				
XX	QY	61 AsnAlaAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly				

DB: 20 Gaps: 1

US-09-985-689A-1 (1-434) x AAV82382 (1-3003)

QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 |||||
 Db 1470 AATGACGTGCCCGTGGCATGTGTAAAGCAGACGTCGCACAAAATAACTTTGGCTATAT 1529
 QY 21 GlyGlnGlyClnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 |||||
 Db 1530 GGACAAGGACAGATTGTGACAGTTGCTGATCTGGGCTTGATACAGGAAGAAATGCACAT 1589
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 |||||
 Db 1590 TCGATGCAATGAAGCATTTCCCGGTAAAGTTACCCACCTATATGCACGGGAGAACGAAT 1649
 QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 |||||
 Db 1650 AACGCCAATCATCCAATGGACATGGNACCATGTTGCTGGATCTGTGTAGGAAT --- 1709
 QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 |||||
 Db 1707 GCTACAAATAAAGGATGGCACCGCAAGCCAATCTAGCTTTCAATCTATTATGGATAGT 1766
 QY 101 GlyGlyGlyLeuGlyClyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 |||||
 Db 1767 GGTGGAGGCTGGGAGGACTACTGCTAATCTACAAACATATTTCAGTCAAGCATATAGT 1826
 QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140
 |||||
 Db 1827 GCTGAGCGAGAAATTCATACGAATTCATGGGGGGCTCCAGTAACGGTGGCTATACGACA 1886
 QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 |||||
 Db 1887 CACTCTCGAAATGTTGATGATTATGTGAGAAAAAATCATATGACGATTCCTTTTTCGGGCC 1946
 QY 161 GlyAsnGluGlyProAsnGlyClyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 |||||
 Db 1947 GGAATGAGGAGCCAGGTAGCCGTAACATCAGTGCACACAGGACAGCAAAATGCCGAT 2005
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 |||||
 Db 2007 ACAGTTGGGGCAACCGAAACCTACGTCCAAGCTTCGGATCTTATGCGGATATATTAAAC 2066
 QY 201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 |||||
 Db 2067 CATGTTGCTCAATTCTCTTCACGAGGTCCCTACTAGAGATGGACGTTATTAAGCCGAGCGTC 2126
 QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 |||||
 Db 2127 ATGGCACACGTACGTATATTCTCTCTGCTAGATCATCATAGCTCCAGATTCCTCAATC 2186
 QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 |||||
 Db 2187 TGGGCAACCATGATAGTAAATATGCCATACATGGGTGCTTCTATGGCTACTCCAAAT 2246
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 |||||
 Db 2247 GTACGAGTAATGTTGCACAATTAAGGAGGCAATTTGTGAAAAATAAGGGGTAACTCCT 2306
 QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 |||||
 Db 2307 AAGCCTTCCCTTTTAAAGCTGCTTTAATTGACGCTGCTCGGATCTTGGACTTGGCTTT 2366
 QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 |||||
 Db 2367 CCAATGCTTAACCAAGGATGGGAGAGTAACGTTAGATAAAATCCCTAAATGTGCGCATTT 2426
 QY 321 ValAsnGluSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 |||||
 Db 2427 GTCAATCAACAGGAGCCCTTTATCAACAACTCAAAGCAACATATTCGTTTACGGCTCAA 2486

Accession	Protein	Sequence	Position
Db		2367 CCAATGCTAACCAAGGATGGGAAGAGTAACGTTAGATAAAATCCCTAAATGTGCATTT	2426
QY	ValAsnGluSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr	321	340
Db		2427 GTCAATGAACAGAGCCCTTTATCAACAACTCAAAAGGCAACATATTCGTTTTCAGGCTCAA	2486

```

Db      2487 GCTGGTAAACCCCTTAAATAATACACTTGTGGTCAGATGCCACAGGTAGCACGCGCA 2546
      361 SerValThrLeuValAsnAspLeuValThrAlaProAsnGlyThrGlnTyr 380
      2547 TCACTAACTTAGTGAATGATTAGACTTACTAATCACTGCACCAAAATGGAACATAAC 2606
      381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGlu 400
      2607 GTCCGGAATGACTTTACAGCACCGCTATGATACAAATGGGATGGCAGAAACACGCTGGAA 2666
      401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
      2667 AATGTGTTTATCAATGCTCTCAAGCGAAGCTATACAGTCGAAGTCGAGCTTACAT 2726
      421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
      2727 GTACCACTAGTCGCGCAACCTTTTCTTTAGCGATTGTACAT 2768

```

RESULT 5

AAQ27516 standard; DNA; 1299 BP.

XX AAQ27516;

XX 05-FEB-1993 (first entry)

XX Alkali-protease Ya enzyme gene.

XX Alkali resistance; surface active agent resistance;

XX detergency improver; ss.

XX Bacillus sp. Y.

XX Key Location/Qualifiers

XX CDS 1..1299

XX /*tag= a

XX JP04197182-A.

XX 16-JUL-1992.

XX 28-NOV-1990; 90JP-0327110.

XX 28-NOV-1990; 90JP-0327110.

XX (LLOY) LION CORP.

XX WPI; 1992-288440/35.

XX P-PSDB; AAR26274.

XX DNA coding alkali-protease Ya enzyme - has good alkali and

XX surfactant resistance and improves detergency

XX Claim 3; Page 2; 17pp; Japanese.

XX The sequence is that of the alkali-protease Ya enzyme gene which can

XX be used in the recombinant production of Ya enzyme. Ya enzyme is

XX excellent in alkali resistance and surface active agent resistance

XX and improves detergency.

XX SQ Sequence 1299 BP; 401 A; 246 C; 297 G; 355 T; 0 other;

Alignment Scores:

Pred. No.: 3,27e-144 Length: 1299

Score: 1986.50 Matches: 379

Percent Similarity: 94.01% Conservative: 29

Best Local Similarity: 87.33% Mismatches: 25

Query Match: 88.41% Indels: 1

DB: 13 Gaps: 1

US-09-985-689A-1 (1-434) x AAQ27516 (1-1299)

```

QY 1 AsnAspValAlaIleValGlyLeuValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
      1 AATGATGTAGCAAGAGGATAGTAAAGCTGATGTGCACAAACAATACGGATTATAT 60
      21 GlyGlnGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
      61 GCACAAGGTCACACTAGTTGCAGTAGCGGACACAGGCTTAGATACAGGTCCTACGATAGT 120
      41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
      121 TCTATGCTGAAGCAATCCCGCGGAAATACACAGCTCTTTACCGTTAGAGAGAACTAAT 180
      61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
      181 AATGGAGTGATCCGAATGGGCATGGCACACATGTAGCAGGTTCTGTACTTGGTAAT 237
      81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
      238 GCTTTAAATAAAGGAATGGCTCCGCAAGCTAAGTTAGTCTTCAATCTATTATGATAGC 297
      101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
      298 AGCGGAGGATTAGTGGCTTACCATCGAACTTAATACGTTATTAGTCAAGCTTGAAT 357
      121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAlaAsnGlyAlaTyrThr 140
      358 GCTGGAGCAAGAATTCATCTACTCTTGGGAGCCCCCAGCAATAATGAGCGTACACTGCT 417
      141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
      418 AACTCGAGCAAGTGGATGATGATGTTTCGAAATAATGATATGACGGTACTTTTGCAGCT 477
      161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
      478 GGTAATGAAGGTCCTAATTCAGAGACAATAGTCTCCAGGTACAGCGCAAAATGCTATT 537
      181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
      538 ACGGTGCGGCGCAACGGAAACTATCGCCCAAGCTTCGGTTCGATAGCAGATAACCCAAAT 597
      201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
      598 CATATTGCACAATTTTCATCGAGAGGAGCTACGAGGATGGACGAAATTAAGCCTGACGTA 657
      221 MetAlaProGlyThrPheIleLeuSerSerAlaArgSerSerLeuAlaProAspSerPhe 240
      658 ACAGCTCTCTGGACATTTATTTATCAGCAGCTTCTTCCCTTAGCTCCAGACTCTTCGTTT 717
      241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
      718 TGGCGCAATTATACAGATAAATACGCGTATATGGCGGTATATGGCGACACCTATT 777
      261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
      778 GTTCAGGGAATTCGCGCAATTACGTGAGCATTTTATAAAAAATAGAGGTATTACTCT 837
      281 LysProSerLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
      838 AAGCCTCTTTAATAAAGCTGCATCTATCCGTGCTACTGATGTTGTTTAGATAT 897
      301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
      898 CCTAGTGGTGACCAAGGCTGGGGCGGTGTTACTCTAGATAAATCTTAATATAGCGTAT 957
      321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
      958 GTCATGAAGCAACTGCTATAGCCACGAGCAAAAGCAACGATATTCGTTCACAGCACAA 1017
      341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
      1018 GCGGGTAAACCTTTAAAAATCTCGTTAGTATGGACAGATGCTCTCTGGAAGTACAACTGCA 1077
      361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380

```

Db 1078 TCTTATACACTAGTCTTAATGATTTAGATCTAGTTATTACTGCTCGAATGGACAAATAT 1137
 QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
 Db 1138 GTAGGAATGATTTTATGTTATCTTATGATAAATACTGGATGGTTCGCAACAATGTTGAG 1197
 QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrThrThrIleGluValGlnAlaTyrAsn 420
 Db 1198 AACGTTATTATAAAGCTCCGCAATCTGGACGTATATAATTGAGGTTCAAGCGTATAAT 1257
 QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1258 GTACCATCTGGCCACAGCGTTTCTCACTAGTCTGCTACAT 1299

RESULT 6

AAT85667

ID AAT85667 standard; DNA; 1977 BP.

XX AAT85667;

XX 20-APR-1998 (first entry)

XX Thermococcus protease coding sequence.

XX Protease; research reagent; thermal stability; thermococcus celer; ss.

XX Thermococcus celer DSM-2476.

XX WO9721823-A1.

XX 19-JUN-1997.

XX 07-NOV-1996; 96WO-JP03253.

XX 12-DEC-1995; 95JP-0323285.

XX (TAKI) TAKARA SHUZO CO LTD.

XX Asada K, Kato I, Mitta M, Morishita M, Takakura H;

XX Tsunawasa S, Yamamoto K;

XX WPI; 1997-332794/30.

XX P-PSDB; AAW24121.

Protease(s) and genes encoding them obtained from Thermococcus and
 Pyrococcus strains - have extremely high thermal stability and are
 useful industrially and as research reagents

PS Claim 3; Page 86-87; 159pp; Japanese.

CC This sequence represents the coding sequence for the protease from
 CC Thermococcus celer DSM-2476. This sequence encodes a protease of the
 CC invention. The proteases of the invention have extremely high thermal
 CC stability. The proteases can be used as research reagents, and
 CC industrially in the food, drug and chemical industries.

XX Sequence 1977 BP; 453 A; 659 C; 554 G; 311 T; 0 other;

Alignment Scores:

Pred. No.:	1.14e-25	Length:	1977
Score:	452.50	Matches:	138
Percent Similarity:	44.66%	Conservative:	67
Best Local Similarity:	30.07%	Mismatches:	153
Query Match:	20.14%	Indels:	101
DB:	18	Gaps:	18

US-09-985-689a-1 (1-434) x AAT85667 (1-1977)

QY 8 ValLysAlaAspValAlaGlnSerTyrGlyLeuTyrGlyGlnGlnIleValAla 27
 Db 433 ATAGGGGGCGGATACCGTCTGGAACTCCCTCGCTACGACGGAAGCGGTGGTGGTGGC 492

QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
 Db 493 ATCGTCGATACAGGGTATACAGCGGAAC-----CACCCGGATCTCGAAG 534
 QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn 66
 Db 535 GGCAAGGTCATAGGCTGGTGGTACGACGCGCTCAACGCGAGGTCTGACCCCTACGATGACCAG 594
 QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLys-----84
 Db 595 GGACACGGAACCCAGCTTCGGGTATCTGTCGGGAACCGCGACGCTTAACCTCCAGTAC 654
 QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 101
 Db 655 ATAGCGCTCGCCCGCGGGAAGCTCGTCCGCGCTCAAGGTTCTCGGTGCGCGAGTTCG 714
 QY 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 121
 Db 715 GGAACGCTCTCCACCATCATCGGGGTGTGACTGGGTCTGCTCAGAACAGGACAAGTAC 774
 QY 122 GlyAlaArgIle-----HisThrAsnSer 129
 Db 775 GGGATAAGGGTCAATCAACCTCTCCCTCGGCTCTCCAGAGCTCCGACGACCGACTCC 834
 QY 130 TrpGlyAlaAlaValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrVal 149
 Db 835 CTCAGTCAGCGCGCTCAACACGCTGGGACGCC-----867
 QY 150 ArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThr 169
 Db 868 -----GGTATAGTAGTCTGCTCGCGCGCGGACAGCGGCGCGGACACCTACACC 918
 QY 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg 189
 Db 919 GTCGGCTCACCCCGCGCGGAGGAGGTCATACCGTCGGTCA-----963
 QY 190 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 209
 Db 964 -----GTTGACAGCAACGACACATCGCCAGCTTCTCCAGAGGGA 1005
 QY 210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSer 229
 Db 1006 CCGACCGCGGACGGAAGGCTCAACGCGGAGTCTGCGCCCGCGGCTTGACATCATAGCC 1065
 QY 230 AlaArgSerSerLeuAlaProAspSerSerPheThrAlaAsnHisAspSerLysTyrAla 249
 Db 1066 CCGCGCGCCAGC-----GGAACGAGCATGGGACCCCGATTAACGACTACTACACC 1116
 QY 250 TyrMetGlyGlyThrSerMetAlaThrProfileValAlaGly---AsnValAlaGlnLeu 268
 Db 1117 AAGGCTCTGGAACGACGATGGCCACCCCGACGCTTTCGGGCGTGGCGGCTCATCTC 1176
 QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLys 286
 Db 1177 CAGGCCAC-----CCGAGCTGGACCCCGCAGAGGTGAAG 1212
 QY 287 AlaAlaLeuIleAlaGlyAla-----AlaAspIleGlyLeu 298
 Db 1213 ACCGCTCTATCGAGACCGCGACATAGTCCGCCCAAGAGATAGCGGACATCGCTAC 1272
 QY 299 GlyTyrProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeu-----316
 Db 1273 GGTGCG-----GGTAGGTGAACGCTCTACAAGGCCATCAAGTAC 1311
 QY 317 ---AsnValAlaTyrValAlaAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyr 335
 Db 1312 GACGACTAGCCAAAGCTCACCGGCTCCGTCGCGCACAGGAAAGGCGGACCGAC 1371
 QY 336 SerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaPro 355
 Db 1372 ACCTTCGAGCTCAGCGCGCCACCTTCGTCGACCGCCACCTCTACTGGGAC-----1422
 QY 356 AlaSerThrThrAlaSerValThrLeuValAsnAspLeuValIleThrAlaPro 375

Db	1423	-----					1461
							1461
Qy	376	AsnGlyThrGlnTyrValGlyAsnAspPheThrSerProTyrAsnAspAspTrpAspGly					395
Db	1462	AACGGGAACGAG---GTTGACTACTCTTACACCGGCTACTAC-----					1500
Qy	396	ArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrThrThrIleGlu					415
Db	1501	-----GGCTCGAGAGGTCGGCTACTACACCGCGGCGGAACTGGAGCGGTCAAG					1554
Qy	416	ValGlnAlaTyrAsnValProValGlyProGlnThrPheSerLeuAlaIleValAsn					434
Db	1555	GTCGTACGTACAAAG-----GGCGGCGGAACTACCGAGGTGACGTGCTCAGC					1602
RESULT 7							
AAX05926							
ID	AAX05926 standard; DNA; 1977 BP.						
XX	AAX05926;						
AC							
DT	06-MAY-1999 (first entry)						
XX							
DE	WO9856926 Seq ID 11.						
XX							
XX	Hyperthermostable; protease; thermophilic; bacterium; subtilisin;						
KW	additive; drug; washing agent; foodstuff; chemical synthesis; ds.						
XX							
OS	Thermococcus celer.						
XX							
PN	WO9856926-A1.						
XX							
PD	17-DEC-1998.						
XX							
PF	04-JUN-1998; 98WO-JP02465.						
XX							
PR	10-JUN-1997; 97JP-0151969.						
XX							
PA	(TAKI) TAKARA SHUZO CO LTD.						
XX							
PI	Asada K, Kato I, Morishita M, Shimojo T, Takakura H;						
XX							
DR	WPI; 1999-080907/07.						
DR	P-PSDB; AAW94840.						
XX							
PT	Recombinant hyperthermostable protease from <i>Pyrococcus furiosus</i> -						
PT	and gene encoding it, for large scale production of the protease for						
PT	industrial use.						
XX							
PS	Disclosure; Page 53-54; 82pp; Japanese.						
XX							
CC	The invention relates to a hyperthermostable protease derived from a						
CC	thermophilic bacterium (especially <i>Pyrococcus furiosus</i>). The protease has						
CC	working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10						
CC	(optimum 6-8), and retains more than 90% of its activity after 8 hours						
CC	at 95 deg.C. The invention also provides gene sequences encoding a						
CC	polypeptide of the formula SIG-Ala-Gly-Asn-PRO, where SIG is a signal						
CC	peptide from subtilisin, and PRO is the above protease. Host cells						
CC	(especially <i>Bacillus</i> strains) transformed with vectors comprising the						
CC	genes are used for the recombinant production of the protease. The						
CC	hyperthermostable protease which can be prepared in quantity suitable						
CC	for industrial use, can be used as an additive for drugs, washing agents						
CC	and foodstuffs and for chemical synthesis.						
XX							
SQ	Sequence 1977 BP; 453 A; 659 C; 554 G; 311 T; 0 other;						
Alignment Scores:							
Pred. No.:				Length:		1977	
Score:	1.14e-25			Matches:		138	
Percent Similarity:	452.50			Conservative:		67	
Best Local Similarity:	44.66%			Mismatch:		153	
Query Match:	30.07%			Indels:		101	
DB:	20.14%			Gaps:		18	
	20						


```

294  -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrPlyArgVa 310
      |||
852  AGATGAATAGCCGATATAGCTAGGTCGA-----GGTAGGGT 890
      |||
310  lThrLeuAspIleSerLeuValAlaTyrValAsnGluSerSerSerLeuSerThrSe 330
      |||
891  TAATGCATACAGGCTATAAAC-----TACGATAACTATGCAAAAGCTAGTGTCTCCTGG 944
      |||
330  rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
      |||
945  ATATGTTGCCAACAAAGCAGCCAACTCACCAGTTCGTTATAGCGGAGCTTCGTCGT 1004
      |||
345  uLysIleSerLeuValTyrPheSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
      |||
1005  AACTGCCACATATATCTACTGGACAATGCCAAT----- 1035
      |||
365  lAsnAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
      |||
1036  -AGCGACCTTGATCTTACCTCTACGATCCCAATGGAACCCAG---GTTGACTACTCTTA 1091
      |||
385  eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
      |||
1092  CACCGCCTACTAT-----GGATTCCGAAAAGGTGTGTTATTA 1127
      |||
405  nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
      |||
1128  CAACCAACTGATGGAACATGGACAATTAAGGTTGTAAAGCTACAGC-----GGAAG 1178
      |||
425  oGlnThrPheSerLeuAlaIleValAsn 434
      |||
1179  TCCAACTATCAAGTAGATGTGTAAGT 1206

RESULT 9
AAT85668
ID AAT85668 standard; DNA: 1566 BP.
AC AAT85668;
DT 20-APR-1998 (first entry)
DE Pyrococcus furiosus protease coding sequence.
KW Protease; research reagent; thermal stability; pyrococcus furiosus; ss.
OS Pyrococcus furiosus DSM-3638.
FH Key Location/Qualifiers
FT CDS 1..1566
FT /tag= a
FT /transl_except= (pos: 1282..1284, aa: Xaa)
FT /note= "Xaa= Gly, Val"
XX WO9721823-A1.
XX
XX 19-JUN-1997.
XX
XX 07-NOV-1996; 96WO-JP03253.
XX
XX 12-DEC-1995; 95JP-0323285.
XX
XX (TAKI ) TAKARA SHUZO CO LTD.
XX
XX Asada K, Kato I, Mitta M, Morishita M, Takakura H;
XX Tsunasa S, Yamamoto K;
XX WPI; 1997-332794/30.
XX P-PSDB; AAW24122.
XX
XX Protease(s) and genes encoding them obtained from Thermococcus and
XX Pyrococcus strains - have extremely high thermal stability and are
XX useful industrially and as research reagents
XX
```

```

PS Claim 7; Page 90-91; 159pp; Japanese.
XX This sequence represents the coding sequence for the protease from
CC Pyrococcus furiosus DSM-3638. This sequence encodes a protease of the
CC invention. The proteases of the invention have extremely high thermal
CC stability. The proteases can be used as research reagents, and
CC industrially in the food, drug and chemical industries.
XX Sequence 1566 BP; 467 A; 342 C; 372 G; 384 T; 1 other;

Alignment Scores:
Pred. No.: 5,18e-23 Length: 1566
Score: 416.50 Matches: 141
Percent Similarity: 42.55% Conservative: 59
Best Local Similarity: 30.00% Mismatches: 148
Query Match: 18.54% Indels: 122
DB: 18 Gaps: 20

US-09-985-689A-1 (1-434) * AAT85668 (1-1566)
QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyCl 24
      |||
Db 24 GTCTGCAGCTCAAGTTATGCGCACTTACGTTTGGAACTTGGATATGATGTTCTCGAAT 83
      |||
QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGl 44
      |||
Db 84 CACAATAGGAATAATTGACACTGGAATTGAC-----GCTTCATCCTCC 125
      |||
QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
      |||
Db 126 AGATCTCCAAGGAAAGTA-----ATTGGTGGGTAGATTTTGTCAATG 170
      |||
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGl 78
      |||
Db 171 TAGGAGTTATCCATACGATGACCATGACATGGAACATCATGTAGTTCAATAGCAGCTGG 230
      |||
QY 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
      |||
Db 231 TACTGGAGCAGCAAGTAATGCAAGTACAGGAATGCTCCAGGAGCTTAAGTGGAGTTGA 290
      |||
QY 94 eGlnSerIleMet-----AspSerGlyGlyLeuGlyLeuGlyLeuProSerAsnLeuGl 112
      |||
Db 291 AATTAAGGTTCTAGTCCGATGTTCTGGAAGCATATCTACTATAATTAAGGAGTTGA 350
      |||
QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
      |||
Db 351 GTGGGCGGTTGATAACAAAGATAAGTACGGAATTAAGGTCAATTAATCTTCTCTGTTTC 410
      |||
QY 132 a-----AlaValAsnGlyAlaTyrThrTh 140
      |||
Db 411 AAGCCAGCTCAGATGGTACTGAGCTCTAAGTCAGGCTGTTAATGACAGCTGGGATGC 470
      |||
QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
      |||
Db 471 T-----GGATAGTGTGTTGTTGTTGCCGC 494
      |||
QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIl 180
      |||
Db 495 TGGAAACAGTGGGACCTTAACAGTATACATCCGTTCTCCAGCAGCTGCACAAAGTTAT 554
      |||
QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
      |||
Db 555 TACAGTTGGAGCC-----GTTGACAGATGATGA 581
      |||
QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
      |||
Db 582 TGTATATAACAAAGCTTCTCAAGCAGAGGGCCAACTGCAGACGGCAGGCTTAAGCTGAGGT 641
      |||
QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPh 240
      |||
Db 642 TGTGTCTCCAGGAACCTGGATAATTGCTCCAGAGCAAGT-----GGAAGTACAT 692
      |||
QY 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIl 260
      |||
```

Db 693 GGTCAACCAATTAATGACTATTACACAGCAGCTCTGGGACATCAATGGCACTCTCA 752
 QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
 Db 753 CGTAGCTGGTATTGCGACCTCTTGCTCAA-----GCACACCC 791
 QY 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
 Db 792 GAGCTGAGCTCCAGCAAAAGTAAACAGCCCTCATAGAACTGCTGATATCGPAAAGCC 851
 QY 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgVa 310
 Db 852 AGATGAATATAGCGATATAGCTACGCTGCA-----GGTAGGGT 890
 QY 310 lThrLeuAspLysSerLeuAsnValAlaLysValAlaLysValAlaLysSerSerLeuSerThrSe 330
 Db 891 TAATGCATACAGGCTATAAC-----TACGATAACTATGCAAGCTAGTGTCTACTGG 944
 QY 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
 Db 945 ATATGTGCGCAACAAAGGCGACCAAACTCACCAGTTCGTTATTAGCGGAGCTTCGTCGT 1004
 QY 345 uLysIleSerLeuValTyrSerAspAlaProAlaSerThrAlaSerValThrLeuVa 365
 Db 1005 AACTGCCACATTATAGTGGCAATGCCAAT----- 1035
 QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
 Db 1036 -AGCGACCTTGATCTTACCTCTACCATCCCAATGGAACCCAG----GTTGACTACTCTTA 1091
 QY 385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
 Db 1092 CACCCCTACTAT-----GGATTGCAAAAGTTGGTTATTA 1127
 QY 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
 Db 1128 CAACCCACTGATGGAACATGGACAAATTAAAGTTGTAGCTACAGC-----GGAAG 1178
 QY 425 oGlnThrPheSerLeuAlaIleValAsn 434
 Db 1179 TGCAACTATCAAGTAGATGTGGTAAGT 1206

RESULT 10

AAT85695

ID AAT85695 standard; DNA; 1962 BP.

XX AC AAT85695;

XX DT 20-APR-1998 (first entry)

XX DE Pyrococcus furiosus PFUS protease coding sequence.

XX KW Protease; research reagent; thermal stability; pyrococcus furiosus; ss.

XX OS Pyrococcus furiosus DSM-3638.

XX PN WO9712823-A1.

XX PD 19-JUN-1997.

XX PF 07-NOV-1996; 96WO-JP03253.

XX PR 12-DEC-1995; 95JP-0323285.

XX XX (TAKI) TAKARA SHUZO CO LTD.

XX XX Asada K, Kato I, Mita M, Morishita M, Takakura H;

XX XX Tsunashawa S, Yamamoto K;

XX XX WPI; 1997-332794/30.

XX DR P-PSDB; AAW24129.

XX XX

PT

Protease(s) and genes encoding them obtained from Thermococcus and
 Pyrococcus strains - have extremely high thermal stability and are
 useful industrially and as research reagents

Disclosure; Page 123-125; 159pp; Japanese.

PS

XX

CC

CC

CC

CC

CC

CC

SQ

This sequence represents the coding sequence for the protease from
 Pyrococcus furiosus DSM-3638. This sequence encodes a protease of the
 invention. The proteases of the invention have extremely high thermal
 stability. The proteases can be used as research reagents, and
 industrially in the food, drug and chemical industries.

Sequence 1962 BP; 602 A; 399 C; 471 G; 490 T; 0 other;

Alignment Scores:

Pred. No.:	6,77e-23	Length:	1962
Score:	416.50	Matches:	141
Percent Similarity:	42.55%	Conservative:	59
Best Local Similarity:	30.00%	Mismatches:	148
Query Match:	18.54%	Indels:	122
DB:	18	Gaps:	20

US-09-985-689a-1 (1-434) x AAT85695 (1-1962)

QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGly 24

Db 420 GTCTGCAGCTCAAGTTATGGCAACTTACGTTTGGAACTTGGGATATGATGGTTCTGGAAT 479

QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG 44

Db 480 CACAATAGGAATAATTGACACTGGAATTGAC-----GCTTCTCATCC 521

QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63

Db 522 AGATCTCCACAGAAAGTA-----ATTGGGTGGTAGATTTCCTCATGC 566

QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG 78

Db 567 TAGGAGTTATCCATACGATGACCATGGACATGGAACCTCATGTAGTTCAATAGCAGCTGG 626

QY 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94

Db 627 TACTGGAGCAGCAAGTAATGCAAGTACACAGGAATGCTCCAGGAGCTAAGCTGCGGG 686

QY 94 eGlnSerIleMet-----AspSerGlyGlyLeuGlyLeuGlyLeuProSerAsnLeuG 112

Db 687 AATTAGGTTCTAGTGCCGATGTTCTTGGAGCATATCTACTATATAATTAGGGAGTTGA 746

QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132

Db 747 GTGGGCGTTGATACCAAGATAAGTACGGAATTAAGGTCATTAATCTTCTTGGTTC 806

QY 132 a-----AlaValAsnGlyAlaTyrThr 140

Db 807 AAGCCAGAGCTCAGATGGTACTACGCTCTAAGTCAGCTGTTAATGCGAGCTGGGATCC 866

QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160

Db 867 T-----GGATTAGTTGTTGGTTCGCGC 890

QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaAl 180

Db 891 TGCAACAGTGGACCTAACCAAGTATACATCGGTTCTCCAGCAGCTGCAACCAAGTTAT 950

QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200

Db 951 TACAGTTGGAGCC-----GTTGACAAAGTAGTA 977

QY 200 nHisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220

Db 978 TGTATACAAAGCTTCTCAAGCAGAGGCGCACTGCGAGCGGCGGCTTAAGCTGAGT 1037

QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPh 240

RESULT 11

RESOL 11
AAX05929
ID AAX05929 standard: DNA: 1962 BP.

XX AC AAX05929:

XX
DT
06-MAY-1999 (first entry)

XX Hypert hermostable protease encoding DNA.

XX Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
KW additive; drug; washing agent; foodstuff; chemical synthesis; ds.
XX

XX

0	0
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99

XX 04 TUN-1008. 09510-7002465

XX
10 MAY 1967 0710-0151060

WPI; 1999-080907/07.
P-PSDB; AAW94841.

Recombinant hyperthermostable protease from *Pyrococcus* and gene encoding it, for large scale production of the industrial use.

Disclosure: Page 59-60; 82pp; Japanese.

The invention relates to a hyperthermostable protease (thermophilic bacterium (especially *Pyrococcus furiosus*) working temperature 40-110 deg C (optimum 80-95 deg C) (optimum 6-8), and retains more than 90% of its activity at 95 deg.C). The invention also provides gene sequences (polypeptide of formula SIG-Ala-Gly-Gly-Asn-PRO, where: peptide from subtilisin, and PRO is the above protease (especially *Bacillus* strains) transformed with the vectors (genes are used for the recombinant production of the hyperthermostable protease which can be prepared in quantity for industrial use, can be used as an additive for drugs and foodstuffs and for chemical synthesis.

Sequence 1962 BP; 602 A; 399 C; 471 G; 490 T; 0 other;

Alignment Scores:	6.77e-23	Length:	1962
Pred. No.:	416.50	Matches:	141
Score:	42.55%	Conservative:	59
Percent Similarity:	30.00%	Mismatches:	148
Best Local Similarity:	30.00%	Indels:	122
Query Match:	18.54%	Gaps:	20
DB:	20		

US-09-985-689A-1 (1-434) x AAX05929 (1-1962)

12	ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1	24
QY		
Db		
420	GTCTGCAGCTCAAGTTATGGCAACTTACGTTTGGAACTGGGATATATGATGGTCTCGAAT	479
QY		
24	nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1	44
QY		
Db		
480	CACAATAGGAAATAATTGCACACTGGAATTGAC-----GCTTCTCATCC	521
QY		
44	uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn--	63
QY		
Db		
522	AGATCTCCAAGAAAGTA-----ATTGGGTGGGTAGATTTTGTCAATGG	566
QY		
64	-----AspThrAsnGlyHisGlyThrHisValAlaGlySerValIeuG1	78
QY		
Db		
567	TAGGAGTTATCCATACGATACCATGCCATGGACATGGAACTCATGTCTCAATAGCAGCTGG	626
QY		
78	yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh	94
QY		
Db		
627	TACTGGAGCAGCAAGTAATGGCAAGTACACAGGGAATGGCTCCAGGAGCTAAGCTGGCGGG	686
QY		
94	eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyLeuProSerAsnLeuG1	112
QY		
Db		
687	AATTAAGGTTCTAGGTGCGCATGTTCTTGGAAAGCATATCTACTATAATTAAGGGAGTTGA	746
QY		
112	nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl	132
QY		
Db		
747	GTGGGGCGGTTTGATAACAAAGATAGTACGCGGAATTAAAGCTATTAACTTCTCTCTGGTTC	806
QY		
132	a-----AlaValAsnGlyAlaTyrThrTh	140
QY		
Db		
807	AAGCCAGAGCTCAGTGTACTGACGCTCTAAGTCAGGCTGTTAATGCAGCGCTGGGATGC	866
QY		
140	rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl	160
QY		
Db		
867	T-----GGATTAGTTGTGTGGTTCGCCGC	890
QY		
160	aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI1	180
QY		
Db		

Db 891 TGGAAACAGTGGACCTAACAGTATACAAATCGGTCTCCAGCAGCTGCAAGCAAGTTAT 950
 QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
 Db 951 TACATTTGGAGCC-----GTTGACAAGTATGA 977
 QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
 Db 978 TGTATAAACAGCTTCTCAAGCAGAGGCCAACTCGACAGCGCAGGCTTAAGCCTGAGGT 1037
 QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
 Db 1038 TGTGTCTCCAGGAATCGGATAATGTCTGCCAGAGCAAGT-----GGAACATAGCAT 1088
 QY 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIl 260
 Db 1089 GGGTCACCAATTAATGACTATTACACAGCAGCTCTGGGACATCAATGGCACTCTCA 1148
 QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
 Db 1149 CGTAGCTGTATTGCGAGCCCTCTTGCTCCAA-----GCACACCC 1187
 QY 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
 Db 1188 GAGCTGGACTCCAGACAAAGTAAACACAGCCCTCATAGAAACTGCTGATATCGTAAAGCC 1247
 QY 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgVa 310
 Db 1248 AGATGAATAGCCGATATAGCTACGGTCA-----GGTAGGGT 1286
 QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerLeuSerThrSe 330
 Db 1287 TAATGCATACAAAGGCTATAAAC-----TACGATAACTATGCAAGCTAGTGTCACCTGG 1340
 QY 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
 Db 1341 ATATGTTGCCAACAAAGGCGACCAACTCACCAGTTCGTATTATAGCGGAGCTTCGTTCGT 1400
 QY 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
 Db 1401 AACTGCCACATATATAGTGGACAATGCCAAT----- 1431
 QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
 Db 1432 -AGCGACCTTCATCTTACCTCTAGCATCCCAATGGAACCCAG--GTTGACTACTCTTGA 1487
 QY 385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGluAsnValPheIleAs 405
 Db 1488 CACGCGCTACTAT-----GGATTGCAAAAGGTTGGTTATTA 1523
 QY 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
 Db 1524 CAACCCAACTGATGAACATGACAAATTAAGGTTGTAAGCTACAGC-----GGAAG 1574
 QY 425 oGlnThrPheSerLeuAlaIleValAsn 434
 Db 1575 TGCAAACTATCAAGTAGATGTGGTAAGT 1602
 RESULT 12
 AAT85669
 ID AAT85669 standard; DNA; 1977 BP.
 XX
 AC AAT85669;
 XX
 XX 20-APR-1998 (first entry)
 DT
 DE Protease coding sequence.
 XX
 KW Protease; research reagent; thermal stability; ss.
 XX
 OS Synthetic.
 XX
 PN WC9721823-AL.

XX 19-JUN-1997.
 PD XX 07-NOV-1996; 96WO-JP03253.
 PF XX 12-DEC-1995; 95JP-0323285.
 PR (TAKI) TAKARA SHUZO CO LTD.
 PA Asada K, Kato I, Mitta M, Morishita M, Takakura H;
 PI Tsunasawa S, Yamamoto K;
 XX WPI: 1997-332794/30.
 DR P-PSDB; AAW24123.
 XX
 PT Protease(s) and genes encoding them obtained from *Thermococcus* and
 PT *Pyrococcus* strains - have extremely high thermal stability and are
 XX useful industrially and as research reagents
 PS Claim 11; Page 95-97; 159pp; Japanese.
 XX
 CC This sequence represents the coding sequence for a protease of the
 CC invention. The proteases of the invention have extremely high thermal
 CC stability. The proteases can be used as research reagents, and
 CC industrially in the food, drug and chemical industries.
 XX
 SQ Sequence 1977 BP; 527 A; 562 C; 513 G; 375 T; 0 other;
 Alignment Scores:
 Pred. No.: 6,87e-22 Length: 1977
 Score: 403.50 Matches: 134
 Percent Similarity: 43.07% Conservative: 68
 Best local Similarity: 28.57% Mismatches: 147
 Query Match: 17.96% Indels: 120
 DB: 18 Gaps: 20
 US-09-985-689a-1 (1-434) x AAT85669 (1-1977)
 QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
 Db 420 GTCGCGACGCTCAAGTTATGGCAACTTACGTTTGGAACTTGGCATATGATGTTCTGGAAT 479
 QY 24 nIleValAlaValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
 Db 480 CACAATAGGAATAATTGACACTGGAATTGAC-----GCTTCTCATCC 521
 QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
 Db 522 AGATCTCCAAAGGAAAGTA-----ATTGGTGGGTAGATTTTGTCAATGG 566
 QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
 Db 567 TAGGAGTTATCCATACCATGACCATGGACATGGAACCTCATGTAGCTTCAATAGCAGCTGG 626
 QY 78 YAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
 Db 627 TACTGGAGCAGCAAGTAATGCGCAAGTACAGGGAATGGCTCCAGGAGCTAAGCTGGCGGG 686
 QY 94 eGlnSerIleMet-----AspSerGlyGlyLeuGlyLeuGlyLeuProSerAsnLeuG1 112
 Db 687 AATTAAGGTTCTAGGTGCCGATGTTCTGGAAGCATATCTACTATATAATTAAGGGAGTTGA 746
 QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIle----- 125
 Db 747 GTGGCGCGTTGATAACAAGATAAGTACGGAATTAAGGTCAATTAATCTTCTCTTGGTTC 806
 QY 126 -----HisThrAsnSerTrpGlyAlaAlaValAlaAsnGlyAlaTyrThr 140
 Db 807 AAGCCAGAGCTCGGACGGAACCGACTCCCTCAGTCAGCCGCTCAACACCGCTGGGAGCC 866
 QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaA1 160
 Db 867 C-----GGTATAGTACTGCTGCGTCGCGC 890

Job time : 406.933 secs

```
Db 1172 CAGTCCAAAGGCAAGTACAAAGGCGTGCACCGCGCGATCTCTCAACGCAAGGTC 1231
QY 98 MetAspSerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
Db 1232 CTCGAGACTCCGGT-----TTCGGCGACGACTCCGGCATCTCCCGCATCGAGTGG 1285
QY 118 AlaTyrSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAlaValAsnGlyAl 137
Db 1286 CGCGCGCGCGAGCGCGCCACGTCGTCACATCAGCTGGCGCGATGGACACCGGAG 1345
QY 137 aTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLe 157
Db 1346 ACCGACCGCTGGAGCGCGG-GTCGACAAGCTGTCCGCGCAGAGGGGCTCCTGTTCGC 1404
QY 157 uPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLy 177
Db 1405 CATCGCGCGCGCAACAGGGCGCGAG-----TCGATCGGTTCGCCCGGACGCGCGA 1458
QY 177 sAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAs 197
Db 1459 CGCGCGCTCACCGTCGCGCGC-----GTCCA 1485
QY 197 pAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLysAspGlyArgIle 216
Db 1486 CGACAAGACAAAGCTTCGCGGACTTCTCTCCACCGCGCGCCCTCGCGCGACGGCCAT 1545
QY 216 eLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPr 236
Db 1546 CAAGCGGACGTACCGCTCCCGGTGGACATCACGGCGCTCGGCGGAGGGCAACGA 1605
QY 236 oAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMe 256
Db 1606 CATCGCGCAGAGTCCGTGGAGGACCGCGCGGTACATGACCATCTCCGCGACGTCGAT 1665
QY 256 tAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnAr 276
Db 1666 GCGACCGCGCACGTCGCGCGCGCGCGCTCTGAAGCAGCAG----- 1711
QY 276 gGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu-----IleAlaGlyAl 293
Db 1712 -----CACCCCGACTGGACCTCCGCGCACTGAAGGGCGCGCTCACCGGCTC 1758
QY 293 aAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAs 313
Db 1759 CACCAAG---GGCGGCAAGTACACCCCGTTCCGAGCAGGGTTCGGGCGGATCCAGGCGGA 1815
QY 313 pLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSer----- 328
Db 1816 CAAGGGCTCCAGCAGACCGTGTATCGCGCGACCCCGTCTCGGTGAGCTTCGGCGTCCAGCA 1875
QY 329 -----ThrSerGlnLysAlaThrTyrSerPheThrAlaTh 340
Db 1876 GTGGCGGCACACCGACGACGCGGTTCACCAAGCAGCTGACCTACCGCAACCTCGGCAC 1935
QY 340 rAlaGlyLysProLysIleSerLeuValTyrSerAsp----- 353
Db 1936 CCAGGACGTCACGCTGAGCTGAGCTGACCGCGCGCGCGGTTTCGGGCGGATCCAGGCGGA 1995
QY 354 -----AlaProAlaSerThrThrAlase 361
Db 1996 GCGGCGCTTCTTCACGCTGGCGGCCACACGCGTCCGCGCGCGCGCGCGCTC 2055
QY 361 rValThrLeuValAsnAspLeu-----Va 371
Db 2056 CGTCGACATGACCGCGCACACCGCGCTCGCGCGCAGCGGTGGACGGCGCTACTCGGCGTA 2115
QY 371 lleThrAlaProAsnGlyThrGlnTyrVal 381
Db 2116 CGTGTGCGCACCGCGCGCGCGCAGCGGTC 2146
```

Search completed: July 28, 2003, 04:01:59

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: July 28, 2003, 02:45:54 ; Search time 5415.74 Seconds
(without alignments)
3278.366 Million cell updates/sec

Title: US-09-985-689A-1

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVOAYNPVGPQTFLAIVN 434

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPRO.spool/US09985689/runat_25072003_143032_832/app_query.fasta_1.1166
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCT=0 -LOOPEXP=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USPR=US09985689 -CGN_1_1_5962=runat_25072003_143032_832 -NCPU=6 -ICPU=3
-NO_MMAR -LARGEQUERY -NEG_SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htg_mus.*
34: em.htg_pln.*
35: em.htg_fod.*
36: em.htg_mam.*
37: em.htg_vrt.*
38: em.sy.*
39: em.htgo_hum.*
40: em.htgo_mus.*
41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2247	100.0	1923	1	AB051423	AB051423 Bacillus
2	2183	97.2	1920	1	AB046403	AB046403 Bacillus
3	2143	95.4	1302	1	AB046406	AB046406 Bacillus
4	2125.5	94.6	3003	6	AR069954	AR069954 Sequence
5	2125.5	94.6	3003	6	BD062155	BD062155 Nucleic a
6	1998.5	88.9	1299	1	AB046405	AB046405 Bacillus
7	1994.5	88.8	1299	1	AB046402	AB046402 Bacillus
8	1987.5	88.5	1299	1	AB046404	AB046404 Bacillus
9	1986.5	88.4	2218	6	E03808	E03808 DNA encodin
10	632.5	28.1	60006	1	AF268611	AF268611 unculture
11	523.5	23.3	5890	3	DDU60086	DDU60086 Dictyostell
12	511.5	22.8	5772	3	AF466309	AF466309 Dictyoste
13	497	22.1	6115	3	DDU20432	DDU20432 Dictyostell
14	452.5	20.1	1977	6	AR201152	AR201152 Sequence
15	447	19.9	5109	3	AF263455	AF263455 Dictyoste
16	425.5	18.9	126928	2	AC096673	AC096673 Trypanoso
17	416.5	18.5	1236	6	AR201146	AR201146 Sequence
18	416.5	18.5	1962	6	AR201155	AR201155 Sequence
19	416.5	18.5	12452	1	AE010265	AE010265 Pyrococcu
20	406.5	18.1	14376	1	AE013049	AE013049 Thermoana
21	398	17.7	303450	1	SCO939130	SCO939130 Streptomy
22	381	17.0	299850	1	AP004601	AP004601 Oceanobac
23	366	16.3	3348	1	D83672	D83672 Streptomyce
24	363.5	16.2	300425	1	AP005044	AP005044 Streptomy
25	351	15.6	2539	6	AR202322	AR202322 Sequence
26	349	15.5	291000	1	SCO939105	SCO939105 Streptomy
27	347.5	15.5	292550	1	AP001513	AP001513 Bacillus
28	346.5	15.4	302300	1	AP005034	AP005034 Streptomy
29	346	15.4	2809	6	AR202321	AR202321 Sequence
30	346	15.4	3900	1	AB007809	AB007809 Streptomy
31	338	15.0	1239	1	AF305633	AF305633 Thermoana
32	338	15.0	1239	1	AY028704	AY028704 Thermoana
33	338	15.0	135638	1	AF484556	AF484556 Streptomy
34	337	15.0	12295	1	AE013026	AE013026 Thermoana
35	332.5	14.8	1329	6	AX433519	AX433519 Sequence
36	331	14.7	233780	1	BSUR0010	BSUR0010 Bacillus su
37	328.5	14.6	300800	1	SCO939112	SCO939112 Streptomy
38	327.5	14.6	301150	1	AP04602	AP04602 Oceanobac
39	324.5	14.4	6854	1	BSU39230	BSU39230 Bacillus sp
40	320.5	14.3	2731	1	AF015225	AF015225 Thermococ
41	310.5	13.8	2931	1	BACVPR	M76590 B subtilis
42	310.5	13.8	97015	1	BSGNR	X73124 B subtilis
43	310.5	13.8	212150	1	BSUP0020	Z99123 Bacillus su
44	309.5	13.8	1520	1	BSR238598	AJ238598 Bacillus
45	309	13.8	4197	1	AF239672	AF239672 Pyrococcu

ALIGNMENTS

RESULT 1

AB051423
 LOCUS AB051423 1923 bp DNA linear BCT 10-MAY-2002
 DEFINITION Bacillus sp. KP43 PROF gene for protease, complete cds.
 ACCESSION AB051423
 VERSION AB051423.2 GI:20521154
 KEYWORDS
 SOURCE Bacillus sp. KSM-KP43
 ORGANISM Bacillus sp. KSM-KP43
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 1 Itoh, S. and Saeki, K.
 new protease
 TITLE Unpublished
 JOURNAL
 REFERENCE 2 (bases 1 to 1923)
 AUTHORS Saeki, K.
 TITLE Direct Submission
 JOURNAL Submitted (21-NOV-2000) Katsuhisa Saeki, KAO, CORPORATION;
 2606 AKABANE, ICHIKAI MACHI, HAGA, TOCHIGI 321-3486, Japan
 (E-mail: 387185@kasanet.kao.co.jp, Tel: 81285687471 (ex. 7471),
 Fax: 81285687403)
 COMMENT On May 9, 2002 this sequence version replaced gi:14164344.
 FEATURES
 Location/Qualifiers
 1..1923
 /organism="Bacillus sp. KSM-KP43"
 /mol_type="genomic DNA"
 /strain="KP43"
 /db_xref="taxon:109322"
 1..1923
 /gene="PROF"
 1..1923
 /gene="PROF"
 /codon_start=1
 /transl_table=11
 /product="protease"
 /protein_id="BAB53674.2"
 /db_xref="GI:20521155"
 /translation="MRKKKKVFLSVLSAAALSTVALSNPSAGGARNFLDFKGIOTT
 QLEKFGSKGGTGAALFVSENVKLPKGLQKLETPANNKLHIQFNGPILLETQK
 TDEKFGAKILQYIPDYAVIEVEGVKFSATSTIEHVESVEPYLPIYRIDPQLFTKAS
 ELKVALDITKOKKEVOLRGIOIAQFAISNDVLYITAKPEYKVMNDVARGIVKADV
 AQSGLYCGQGVAVADTGLDTRNDSSMHEAFRGKITALVALGRTNNANDNGHGT
 HVAQSLGSLGNTGMAPOANLVFQSIIMDSGGGLGGLPSNLQTLFQASAGARIHTN
 SWGAAYNGAYTINDSRVDYVYKNDMTILFAAGNEGPNGGTISAPGTAKNAITVGPTE
 NLRPSFGSYADNNHVAQFSSRGPTKDRIRKPDVMAQPTFILSARSLAPDSSFVANH
 DSKVAYMGTSMATPIVAGNVAOLREHFVKNRGIITPKSLKALLAGAADIGLVPN
 GNQGWGRTLDKSLNAVYNSSLSSTOKATYSFTATAGKPLKISLVWSDAPASTA
 SVTLNDDLDLITANCTQYVGNDFSTPYNDWNGRNNVFNINAPQSGTYTIEVQA
 YNPVGPOTFSLAIYN"
 BASE COUNT 576 a 418 c 475 g 454 t
 ORIGIN

Alignment Scores:
 Pred. No.: 3,44e-133 Length: 1923
 Score: 2247.00 Matches: 434
 Percent Similarity: 100.00 Conservative: 0
 Best Local Similarity: 100.00 Mismatches: 0
 Query Match: 100.00 Indels: 0
 DB: 1 Gaps: 0

US-09-985-689a-1 (1-434) x AB051423 (1-1923)

Qy 1 AsnAspValAlaArgGlyLeuValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 |||||
 Db 619 AATGATGTCGCGTGAATTTGCAAGGCGATGTGGCTCAGACGACTACGGGTGTGAT 678
 Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 |||||
 Db 679 GGACAAAGGACACATCGCGGTGCGCATACAGGCGCTTGATACAGTCCGAATACAGT 738
 Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 |||||
 Db 739 TCGATGCATGAAGCCCTTCGCGGGAAAAATTTACTGCATTATATGATTGGGACGACGAAT 798

Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 |||||
 Db 799 AATCCCAATGATACGAATGGTCATGGTAGCATGTGGCTGCCGTATATTAGGAACGCG 858
 Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 |||||
 Db 859 TCCACTAATAAAGGAATGGCGCTCAGGCGAATCTAGCTTCCCAATCTATCATGGATAGC 918
 Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 |||||
 Db 919 GGTGGGGACTTGGAGGACTACCTTCGAATCTGCAAACTTATTCAGCAAGCATACAGT 978
 Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAlaAsnGlyAlaTyrThr 140
 |||||
 Db 979 GCTGGTCCAGAAATTCATAAACTCCTGGGAGCAGCAGTGAATGGGGCTTACACAACA 1038
 Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 |||||
 Db 1039 GATTCAGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCCTTTTCGCTGCC 1098
 Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 |||||
 Db 1099 GGAATGAAGGACCGAAGCGCGAACCATCATGTCACCGACGACAGCTAAATGCAATA 1158
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 |||||
 Db 1159 ACAGTCGAGCTACGGAACCTCCGCCCAAGCTTTGGGTCTTTATGCGGACAAATATCAAC 1218
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 |||||
 Db 1219 CATGTGGCACAGTTCTTTCACGTGGACCGACCAAGGATGGACGGATCAACACCGGATGC 1278
 Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 |||||
 Db 1279 ATGGCACCGGGAACGTTTCACTATCATCAGCAAGATCTTCTCTTGACCGGATTCCTCCTC 1338
 Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 |||||
 Db 1339 TGGGGCAACCATGACAGTAAATATGATACATGGTGGACGCTCCATGGCTACACCGATC 1398
 Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 |||||
 Db 1399 GTTGTGGAACGTCGCACAGCTTCGTGACCAITTTGTGAAAAACAGAGGCATCACACCA 1458
 Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 |||||
 Db 1459 AAGCCTTCTCTATTAAGAGCGGCACTGATTCGCGGTGCAGCTGACATCGGCTTGGCTAC 1518
 Qy 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 |||||
 Db 1519 CCGAAGCGTAAACCAAGGATGGGACGAGTGACATGGATAAATCCTGGAACGTTGCCCTAT 1578
 Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 |||||
 Db 1579 GTGAAGCAGTCCAGTTCTCTATCCACCAGCAAAAGAGCAGCTACTCGTTTACTGCTACT 1638
 Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
 |||||
 Db 1639 GCCGGCAAGCCTTTGAAAAATCTCCCTGTTATGGTCTGATGCCCTTGGCAGCACACTGCT 1698
 Qy 361 SerValThrLeuValAsnAspLeuAspValIleThrAlaProAsnGlyThrGlnTyr 380
 |||||
 Db 1699 TCCGTAACGCTGCTCAATGATCTGGACCTTGCTCATTCACCGTCCCAATGGCACACAGTAT 1758
 Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
 |||||
 Db 1759 GTAGGAAATGACTTTTACTTCCCAATACATAATGATACTGGGATGGCGCAATACCGTAGAA 1818
 Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 |||||
 Db 1819 AATGATATTATTATGACCCACAAAGCGGGACGCTATACAAATTGAGGTACAGGCTTATAC 1878
 Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434

QY 381 valGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
 |||||.....
 Db 1756 GTCGGGAATGACTTCTCAGCACCAATTTGACAAATACTGGGATGCGCGCAATAACGTAGAA 1815
 QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 |||||.....
 Db 1816 AATGATTATTAATTCGCCCAAGTGGACATATACCAITGAGTGCAAGCATATAAT 1875
 QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 |||||.....
 Db 1876 GTCCCGGTGGACACAAAACCTTCTCGTTGCAATTTGTGAAC 1917

RESULT 3

AB046406 1302 bp DNA linear BCT 23-JAN-2001
 LOCUS Bacillus sp. NV1 PROE gene for protease, partial cds.
 DEFINITION
 AB046406
 ACCESSION
 VERSION
 AB046406.1 GI:12381944

KEYWORDS

SOURCE

ORGANISM

Bacillus sp. NV1
 Bacillus sp. NV1
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

REFERENCE

1 (sites)
 Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and
 Horikoshi, K.

AUTHORS

Novel oxidatively stable subtilisin-like serine proteases from
 alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
 evolutionary relationships

TITLE

Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)

JOURNAL

MEDLINE

PUBMED

11118284

REFERENCE

2 (bases 1 to 1302)
 Saeki, K.
 Direct Submission
 Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
 Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,
 Tochigi 321-3497, Japan (E-mail:387185@kasanet.kao.co.jp,
 Tel:81-285-68-7400, Fax:81-285-68-7403)
 Location/Qualifiers

FEATURES

source

1..1302

/organism="Bacillus sp. NV1"

/mol_type="genomic DNA"

/strain="NV1"

/db_xref="taxon:133781"

1..1302

/gene="PROE"

<1..>1302

/gene="PROE"

/codon_start=1

/transl_table=11

/product="protease"

/protein_id="BAB21269.1"

/db_xref="GI:12381945"

/translation="NDVARGIVKADVAQSSYGLYGQGVVAVDVTGLDRCRNDSSMHE

AFRCKITAIYALGRNTNNDPNHGTHVAGSYLNGTSGKMGAPQANLVFOSVMSDNG

GLGPGNSVTLFQAYSGARIHTNSMGAPVNGAYTDSRNVDVYRKNDMAVLFAA

GNEPNGTISAPSGTAKNALTIVGATENLRPSGSYADINLHVAFSSRQPTKDRILKP

DVMAPGTLLSAPDSLPDSFWANHDSKRYATMGTSMAPIVAGNVQAQRLEFINR

GITPKSLKALILAGVADTDIGLYPSGNOGKRVTLDKSLNFAVNETSSLSTNQKAT

YFTAQSGKPLKISLVSDAPASTASVTLDLIVITAPNGTKYVGNDFTPAYDNN

WDGRNNVNVFINAPQSGTYTVEVQAYNVPGQPAFLAIVN"

BASE COUNT 377 a 283 c 322 g 320 t

ORIGIN

1302

Length:

Matches: 406

Score: 2143.00

Percent Similarity: 97.93%

Best Local Similarity: 93.55%

Query Match: 95.37%

DB: 1

Indels: 0

Gaps: 0

Conservative: 19

Mismatches: 9

Alignment Scores:

Pred. No.: 1302

US-09-985-689A-1 (1-434) x AB046406 (1-1302)

QY 1 AsnAspValAlaArGlyIleValLysAlaAspValAlaGlnSerTyrGlyLeuTyr 20
 |||||.....
 Db 1 AACCATGTTCAGAGGCAATGTAAGCCGATGTTGCCAGAGCAGCATATGGTTTATAT 60
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 |||||.....
 Db 61 GGGCAAGGCAAGTGGTTGCAGTACGGGACTCGGATCGGATACAGGGCGTAATGACAGC 120
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 |||||.....
 Db 121 TCGATGATGAAGCGTTCGAGGAAAGATTACCGCGATATATCCCTTGGAAAGCAAAAC 180
 QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 |||||.....
 Db 181 AACCCCAATGATCCAAACGGACACGGGACGATGTTGCCGAATCTGTTTATGGAACCGT 240
 QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 |||||.....
 Db 241 ACTTCGAATAAAGGATGGCTCCACAAGCTAACTTAGTTTCCCAATCTGTTATGGACAGC 300
 QY 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 |||||.....
 Db 301 AATGCGGACTTGGCGGACTGCCTTCCCAATGTAAGTACATATTACCGCAGGCATATAGT 360
 QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140
 |||||.....
 Db 361 GCCGTGTCAGAAATCCATACGAACCTCATGGGAGCGCTGTAACAGGAGCGTACACTACT 420
 QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 |||||.....
 Db 421 GATTCAGAAACGTAGACGATATGTCGTAATAATGATATGCGGTTCTTTTTCGACCG 480
 QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 |||||.....
 Db 481 GGTACGAAGGSCCGAAATGGCGGCAACATCAGTCTCTCGTACCGCGAAGATGCTATC 540
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 |||||.....
 Db 541 ACAGTAGGGGCAACAGAAACCTGCGCCCAAGCTTTGGATCTTATGCTGCAACATCAAT 600
 QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 |||||.....
 Db 601 CATGTAGCACAGCTTTCTCCCGCGACCTACAAAGGATGACGCTATCAAAACCGGACGTA 660
 QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 |||||.....
 Db 661 ATGGCACCAGGAACATTTATTTATCGGCAAGATCTCTTTGGCTCCGGACTCCTCATTC 720
 QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 |||||.....
 Db 721 TGGGCAACCATGACAGCAAAATATGCTTATATGGGTGAACATCCATGGCGACACCGATT 780
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 |||||.....
 Db 781 GTAGCTGTGTAACGTTGACAGTGTACGTAACATTTTCATCAAAACAGAGGAATCACTCT 840
 QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 |||||.....
 Db 841 AAACCATCTCTGCTGAAGACAGCTCTTATTCGCGGAGCACTGATATCGGCTTGGCTAT 900
 QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 |||||.....
 Db 901 CCGAGTGGAAACCAAGGATGGGGAAGAGTAACATTTGGACAAGTCACTTAATGTAGCTTTC 960
 QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 |||||.....
 Db 961 GTAAATGAACAAAGCTCTTTATCTACTTAAACCAAGGCTAGTATTCTATTACTGCACAA 1020
 QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
 |||||.....
 Db 1021 TCAGGCAACCTTTGAAGATTTTCATTTGTTGTTGCTGATGCAACCGGCAAGTACTTCCGCA 1080

```

QY 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 1081 TCGGTTACATTCGTGAATGATCTCGATCTCGTGATTACAGCTCCAAATGGACAAGTAT 1140
QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
Db 1141 GTTGGAAACGACTTTACTGCTCCCTATGATAATAAATACGTTGGATGGACGTAACAATGTAGAG 1200
QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrThrIleGluValGlnAlaTyrAsn 420
Db 1201 AACCTGTTTATCAATGCTCGCAAGCGGAACGTATACAGTTGAGCTACAGGCTTACAAT 1260
QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1261 GTTCCACAAGGCGCGCAGGCGTTTCTTTGGCTATTGTGAAC 1302

RESULT 4
AR069954
LOCUS AR069954 3003 bp DNA linear PAT 18-FEB-2000
DEFINITION Sequence 41 from patent US 5891701.
ACCESSION AR069954
VERSION AR069954.1 GI:7220842
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 3003)
AUTHORS Sloma, A. and Christianson, L.
TITLE Nucleic acids encoding a polypeptide having protease activity
JOURNAL Patent: US 5891701-A 41 06-APR-1999;
FEATURES
Location/Qualifiers
1..3003
/organism="unknown"
BASE COUNT 982 a 504 c 645 g 872 t
ORIGIN

Alignment Scores:
Pred. No.: 2,86e-125 Length: 3003
Score: 2125.50 Matches: 406
Percent Similarity: 97.93% Conservative: 19
Best Local Similarity: 93.53% Mismatches: 8
Query Match: 94.59% Indels: 1
DB: 6 Gaps: 1

US-09-985-689A-1 (1-434) x AR069954 (1-3003)

QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 1470 AATGACGTGGCCCGTGGCATTGTGAAAGCAGACGTCGCACAAAATAACTTTGGCTTATAT 1529
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1530 GCACAAGACAGATTGTAGCAGTTGCTGATACCTGGGCTTGATACAGGAAGAATGACAGT 1589
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 1590 TCGATGTCATGAAGCATTCGCGGTAAAGATTACCGCACTATATGCACTGGGCAGACGAAT 1649
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 1650 AACGCCAATGATCAAAATGGACATGGAAACCCATGTTGTGGATCTGTGTAGGAAT --- 1706
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 1707 GCTACAATAAAGGATGGCACCAGCCCAATCTAGCTTTTCAATCTATTATGATAGT 1766
QY 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 1767 GGTGGAGGCTGGGAGGACTACCTGCTAATCTACAAACATTATTTCAGTCAAGCATATAGT 1826
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140
Db 1827 GCTGGACGGAATTCATACCAATTCATGGGGGCTCCAGTAAACGGTGCCTATACGACA 1886

```

```

QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 1887 GACTCTCGAAATGTTGATGATTATGTGAGAAAAAATGATATGACGATTTCTTTTTCGGGCC 1946
QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1947 GGAATGAGGACACAGGTAGCGGTACATCACTGACGACGACGACGACGACGACGACGACG 2006
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 2007 ACAGTTGGGGCAACCCAAACCTACGTCGAAGCTTCGGATCTTATGCGGTAATATTAAC 2066
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 2067 CATGTTGCTCAATTTCTTTCACGAGGCTCTACTAGAGATGGAGCTATTAAAGCGGACGTC 2126
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
Db 2127 ATGGCACCAGGTACGTATATCTCTCTGCTAGATCATCATTAGCTCCAGATTCCTCATTC 2186
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 2187 TGGCAAAACCATGATAGTAAATATGCTACATGGGTGGTACTTCTATGGCTACTCCAATT 2246
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 2247 GTAGCAGGTAAATGTTGCCACAATTAAGGGAGCATTTTGTGAAAAATAGAGGGTAACTCCT 2306
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 2307 AAGCTTCCCTTTTAAAGCTGCTTTAATTGAGGGTGGTGGGATGTTGGACTTGGCTTT 2366
QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 2367 CCAATGCTAAACCAAGGATGGGAAGAGTAGCTTAGATAAAATCCCTAAATGTCGCAATT 2426
QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 2427 GTGAATGAACAGCAGCCCTTTATCAACAAGTCAAAAGCAACATATTCGTTTACGGCTCAA 2486
QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db 2487 GCTGTAAACCCCTTAAAAATATCACTTGTGTCAGATGCACAGGTAGCAGCAGCGCA 2546
QY 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 2547 TCAGTAATCTTAGTAATGATTAGACTTAGTAACTCACTGCACCAAAATGGAACATAATAC 2606
QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
Db 2607 GTCGGAATGACTTTACAGCACCGCTATGATAAACAATTCGGATGGCAGAAACACGUGGAA 2666
QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 2667 AATGTGTTTATCAATGCTCTCAAAAGCGGAACGTATACAGTCAAGAGTCAGGCTTACAAT 2726
QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 2727 GTACCAGTAAGTCGCAAAACCTTTCTTTAGCGATTGTACAT 2768

RESULT 5
BD062155 3003 bp DNA linear PAT 27-AUG-2002
LOCUS BD062155
DEFINITION Nucleic acids encoding a polypeptide having protease activity.
ACCESSION BD062155
VERSION BD062155.1 GI:22607760
KEYWORDS JP 2001514529-A/39.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 3003)

```


source

```

1. .1299
/organism="Bacillus sp. SD521"
/mol_type="genomic DNA"
/strain="SD521"
/db_xref="taxon:133780"
1. .1299
/gene="PROD"
<1. .>1299
/gene="PROD"
/codon_start=1
/trans_table=11
/product="protease"
/protein_id="BAB21268.1"
/db_xref="GI:12381943"
/translation="NDVARGIVKADVANNYGLYGGOVAVADTGLDGRNDSSWHE
AFRGKIALYALGETNNANDPNHGTHVAGSVLGNALNKGMAPQNLVPOSIMDSGG
LEGPSNLITLFSQANNAGARLHITNSWGPVNGAYTANSROVDEYVRNNDMTLVFAAG
NCGNSNTLAPGAKNAITVGATENYRPSFGSLADNPNHIAQFSRRGMDTRGRKPD
VTPGTEILSARSLAPDSFWANYNSKYMGTSMTATPIVAGNVAOLREHFIKNRG
ITKPSLIKAALIAAGATDVLGYPSGDQGRVTLQKSLNVAIVNEATALATGQATY
SFOAQAGKPLKLSLWTDAPGSTYSLVNDLIVITAPNGOKYVGNDFSYPDNNW
DGRNVENVFVINAQSGTYTIEVOAYNVPSPQRFSLAIVH"

```

BASE COUNT 409 a 245 c 298 g 347 t

ORIGIN

Alignment Scores:

```

Pred. No.: 1.08e-117 Length: 1299
Score: 1998.50 Matches: 381
Percent Similarity: 94.24% Conservatives: 28
Best Local Similarity: 87.79% Mismatches: 24
Query Match: 88.94% Indels: 1
DB: 1 Gaps: 1

```

US-09-985-689A-1 (1-434) x AB046405 (1-1299)

```

QY 1 AsnAspValAlaAArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
DB 1 AATGATGTAGCAAGAGGAATAGTAAAGCAGACGTTGCACAAAACAATACGGACTATAT 60
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 61 GGACAAGTCAAGTAGTGTGACAGTACGGGTTAGATACAGGTCGTAAACGATAGT 120
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
DB 121 TCTATGATGAAGCATTCGCGGGAATATACAGCTCTTTACGCGTTAGGAAGAACTAAC 180
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
DB 181 AATGCAATGATCCGAATGGGATGGTACGCGATGTAGCAGGTTCTGTACTTGGTAAT --- 237
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
DB 238 GCTTTAAATAAAGGAATGGCTCCCAAGCTAACCTAGCTTCCTCCATCTATTATGGATAGC 297
QY 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
DB 298 AGCGGAGGATAGGTGGATACCATCGAATTTGAATACGTTATTAGTCAAGCTTGAAT 357
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140
DB 358 CTGGGGCTAGAAATTCATACCTCTGTGGGCTGCTCCAGTAAATGGAGCTACACTGCT 417
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
DB 418 AACTCGAGACAAGTGGATAGTAGTTCGAAATATGATATACAGGTAATTTTTCAGCA 477
QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
DB 478 GGTAAATGAAGCTCTTAATTCAGGAACAATAGTCTCCAGGCACAGCGGAAATGCCATT 537
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
DB 181

```

```

DB 538 ACGTCCGCGCCACACGGAAAACTATGCGCCGAGCTTTCGGTTTCATTAGCAGATAACCCAAAT 597
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
DB 598 CATATCGCACAAATTTTCATCAAGAGGAGCTACGAGAGATGGACGAATTAACACGACGTA 657
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
DB 658 ACAGCTCTCTGGACATTTATTTATCAGACAGTTCTCTCTTACGCCACGACCTATA 717
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
DB 718 TGGCGGAATTATACAGTAAGTATGCGTACATGGCGGTACCTCTATGGCAGACCTATA 777
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
DB 778 GTTCGCGGAATGTGCGCAATACGCGAGCATTTTATAAAAAATAGAGGAATTACACCT 837
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
DB 838 AAACCTTCCCTTAATAAAAGCTGCACTTATCCTGGGGCTACTGATGTTGGTTTAGGATAT 897
QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
DB 898 CCAAGTGTGACCAAGGCTGGGGCGTCTTACTCTAGATAAATCGTTAAATGTACGGTAT 957
QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
DB 958 GTCAATGAAGCAACTGCAATAGCAACAGGACAAAAAGCAACGATATTCGTTCCAAGCACAA 1017
QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
DB 1018 CGCGGTAACCTTTAAAAATCTCGTTAGTAGACAGATGACCTGGGAAGTACAACTGCA 1077
QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
DB 1078 TCTATACACTAGTTAATGATTTAGATCTAGTCTAGTTACTGCTCCGAATGGACAAAATAT 1137
QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
DB 1138 GTAGGAATGATTTTATGTTATCTATGATAAATACTGGGATGTTGCGCAACAATGTTGAG 1197
QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
DB 1198 AAGCTATTATATAACGCTCCGCAATCTGGAACTGATACATTAATGAGTTCAACGCTATAAT 1257
QY 421 ValProValGlyProGlnThrPheSerSerLeuAlaIleValAsn 434
DB 1258 GTACCATCTGGCCACACAGCGTTTCTCACTAGCTATCGTACAT 1299

RESULT 7
AB046402
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

```

```

AB046402
Bacillus sp. D6 PROA gene for protease, partial cds.
AB046402.1 GI:12381936
Bacillus sp. D6
Bacillus sp. D6
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and
Horikoshi, K.
Novel oxidatively stable subtilisin-like serine proteases from
evolutionary relationships
Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
20568675
11118284
2 (bases 1 to 1299)
Saeki, K.
Direct Submission
Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,

```

Biological Science Laboratory; Ichikaimachi Akabane 2606, Haga,
Tochigi 321-3497, Japan (E-mail: 387185@ekastanet.kao.co.jp,
Tel: 81-285-68-7400, Fax: 81-285-68-7403)

FEATURES

Location/Qualifiers
1..1299
/organism="Bacillus sp. D6"
/mol_type="genomic DNA"
/strain="D6"
/db_xref="taxon:127889"
1..1299
/gene="PROA"
<1..>1299
/gene="PROA"
/codon_start=1
/transl_table=11
/product="protease"
/protein_id="BAB21265.1"
/db_xref="GI:12381937"
/translation="NDVARGIVKADVAQNNGYGLGQGVAVADTGLDTRNDSSMHE
AFRKITALYALGRITNNANDPNHGHVAGSVLGNALNKGMAPOANLVFQSIMDSSGG
LGGPNSLNTLFSQAWNAGARHTNSWGAPVNGAYTANSROVDEVYRNDMTVLEAAG
NEGPNSTGISAPCTAKNAITVCATENYRPSGSIADPNHIAQFSSRGATROGRKPD
VTAPGTFILSRSLADPSFWANTNSKATYMGTSMTPTVAGNVQAQREHPIKRG
ITPSPKIALIACATDVGLGYPGQDQGRVTLDKSLNVAIVNEATALTTCOKATY
SFQQAQKPLKLSLWTDAGSTTSYTLVNDLIDLITAPNGQKYVGNDFSEYDNNW
DGRNVNVEFTAPQSGTYTIEVQAYNVPSGPQRFSLAIVH"
BASE COUNT 413 a 243 c 298 g 345 t
ORIGIN

Alignment Scores:
Pred. No.: 1.92e-117 Length: 1299
Score: 1994.50 Matches: 380
Percent Similarity: 94.01% Conservative: 28
Best Local Similarity: 87.56% Mismatches: 25
Query Match: 88.76% Indels: 1
DB: 1 Gaps: 1

US-09-985-689A-1 (1-434) x AB046402 (1-1299)

QY 1 AsnAspValAlaAtrGlyLeuValLysAlaAspValAlaGlnSerTyrGlyLeuTyr 20
DB 1 AATGATGAGCAAGAGCAATAGTAAGACAGACGTTGCACAAACAAATACGGACTATAT 60
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 61 GGACAAGTCAGTACGTGACGATAGCGGTTAGATACAGGTCGTAACGATAGT 120
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
DB 121 TCTATGATGAGCAATCCGTCGGGAAATACAGCTCTTTACCGCTTAGGAGAACTAAC 180
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
DB 181 AATGCAATGATCCCAATGGGCATGTACGCATGTAGTGTCTGTGCTTGGTAAT-- 237
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
DB 238 GCTTTAAATAAGGAATGGCTCCGCAAGCTAACTAGTCTTCCAAATCTATTATGGATAGC 297
QY 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
DB 298 ACCGAGGATAGGAGGATACCATCGAATTTAAATACGTTATTAGTCAAGCTTGGAT 357
QY 121 AlaGlyAlaAraGileHisThrAsnSerTyrGlyAlaAlaValAsnGlyAlaTyrThr 140
DB 358 GCTGGCGCTAGAAATTCATACCTCTTGGGGGCCCGACAGTAAATGAGGCGTACACGCT 417
QY 141 AspSerArgAsnValAlaAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
DB 418 AACTCGACACAAAGTGGATGATGTTCGAAACAAATCATATGACGGTATTTTTCAGCT 477
QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180

478 GGAATCAAGACCTAACTCTCGAACAAATAGCGCTCCAGGGACAGCGAAAAATGCCATT 537
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
DB 538 ACGTCGGCGCAACGAAACTACCGCAAGTTTGGTTCATTCAGATAACCTAAT 597
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
DB 598 CATATCGCACAAATTTTCATCGAGAGGAGCTACGAGAGATGGAGCAATTAACACAGCGTA 657
QY 221 MetAlaProGlyThrPheIleLeuSerAlaAraGlySerSerLeuAlaProAspSerPhe 240
DB 658 ACAGCTCTCGAACATTTATATATACGACGCTCTCTTTAGCACGACTCTTCGTT 717
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
DB 718 TGGGCAATTTATACAGTAAGTATGCTATATGGCGGTACCTCTATGGGACACCTATA 777
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
DB 778 GTTGGGGAATGTCGCGCAATTCGCGAGCATTTTATAAAATAGAGCAATACACCT 837
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
DB 838 AAACCTCTCTTAATAAAGCTGCATATCGCTGGGCTACTGATGTTGGTTAGGATAT 897
QY 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
DB 898 CCAAGTGTGACCAAGCTGGGCGTGTACTTTAGATAAATCGTTAAATGTAGCGTAT 957
QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
DB 958 GTCAATGAGCAACTGCATTAACACAGGACAAAGCAACGATTCGTTCCAAACGCAA 1017
QY 341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrThrAla 360
DB 1018 CGCGGTAAACCATTAATAATCTGTTAGTAGACAGATGCACCTCGAAGTACAACGCA 1077
QY 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
DB 1078 TCTTATACACTAGTTAATGATTTAGATCTAGTTATTACTGCTCGCAATGGACAAAAAT 1137
QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
DB 1138 GTAGGTAAATGATTTAGTTATCCTTATGATTAATTTGGATGGTGGCAACAATGTTGAG 1197
QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
DB 1198 AACGTATTATAAACGCTCGCAATCTGGAACGATATACAATTGAGTTCAAGCGTATAAC 1257
QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
DB 1258 GTTCCATCTGGACACAGCGTTTCTCACTAGTATCGTACAT 1299

RESULT 8
AB046404
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

AB046404
Bacillus sp. Y PROC gene for protease, partial cds.
AB046404.1 GI:12381940
Bacillus sp. Y
Bacillus sp. Y
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
1 (sites)
Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and
Horikoshi, K.
Novel oxidatively stable subtilisin-like serine proteases from
alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
evolutionary relationships
Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
20568675
11118284

REFERENCE

2 (bases 1 to 1299)
Saeki, K.
Direct Submission
Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,
Tochigi 321-3497, Japan (E-mail: 387185@kasanet.kao.co.jp,
Tel: 81-285-68-7400, Fax: 81-285-68-7403)
Location/Qualifiers

FEATURES

source
1. .1299
/organism="Bacillus sp. y"
/mol_type="genomic DNA"
/strain="y"

gene

1. .1299
/db_xref="taxon:133779"

CDS

<1..>1299
/gene="PROC"
/gene="PROC"
/codon_start=1
/evidence=experimental
/transl_table=1
/product="protease"
/protein_id="BAB21267.1"
/db_xref="GI:12381941"

/translation="NDVARGIVKADVAQNNYGLYGQVAVADTGLDTRNDSSMHE
AFRGKITALYLCRTNNSADPNHGHTVAGSVLGNALNKGMAPOANLVQSIDSSGG
LGLPLSNLTLSQANAGARIHTNSWAGAYTANSROVDYVRNNDMTVLFAG
NEGPNSTISAPGTAKNAITVGTENYRPSFGSIADPNHIAQFSRKGATRDGRIKPD
VTAPGTFILSARSSLPDPSFWANYSKYVMGTSMTAPIVAGNVAOLREHFIKRG
ITPKPSLKAALITAGATDVGLGYPNGDQGWGRTVFNKSLNVAIVNEATATATGOKATY
SFOAQAGPLKLSLWYTDAPGSTTASYTLVNDLVLITAPNGOKYVGNDFSYPDYNNW
DGRNVENVFNAPSGYILIEVQATNVPFGPQFSLAIVH"

BASE COUNT 404 a 245 c 295 g 355 t

ORIGIN

Alignment Scores:

Pred. No.: 5,33e-117 Length: 1299
Score: 1987.50 Matches: 379
Percent Similarity: 94.01% Conservative: 29
Best Local Similarity: 87.33% Mismatches: 25
Query Match: 88.45% Indels: 1
DB: 1 Gaps: 1

US-09-985-689A-1 (1-434) x AB046404 (1-1299)

Qy 1 AsnAspValAlaAArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 1 AATGATGAGCAAGAGGATAGTAAAGCTGATGTGCACAAACAAATACGGATATAT 60
Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAspSer 40
Db 61 GGACAAGTCAAGTAGTTCAGTAGCGGCACACAGCTTAGATACAGGTCGTAAACGATAGT 120
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 121 TCTATGATGATGAAGCATTCGCGGGAATATCACAGCTCTTACGCGTTAGGAAGACTAAT 180
Qy 61 AsnAlaAspThrAsnGlyHisClyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 181 AATCGAGTAGTCCGATGGCATGGCACATGTAGCAGGTCTCTGTACTTGGTAAT--- 237
Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 238 GCTTTAAATAAAGAAATGGCTCCCAAGCTAACCTAGCTTCCCAATCTATTATGGATAGC 297
Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 298 ACCGGAGGATAGTGGCTTACCATCAACCTAAATACGTTATTTAGTCAAGCTTGGAAAT 357
Qy 121 AlaGlyAlaArgIleHisThrAspSerTrpClyAlaAlaValAlaAsnGlyAlaTyrThr 140
Db 358 GCTGGAGCAGAAATTCATACCTCTTGGGAGCCCAAGTAAATGGAGCGTACACTGCT 417
Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160

Db 418 AACTCGAGACAAGTGGATGAATATGTTGAAATAATGATATGACGGTACATTTTTCAGCT 477
Qy 161 GlyAsnGluClyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 478 GGTAAATGAAGTCCCTAATTCAGAGACAATAGTCTCCAGGTACAGCGAAATGCTATT 537
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 538 ACGGTCGCGCAACGAAACTATCCCAAGCTTCGGITCGATAGCAGATAACCCAAAT 597
Qy 201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 598 CATATGACAAATTTTCATCGAGAGAGCTAGAGGATGGAGCAATTAAGACCTGACGTA 657
Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 658 ACAGCTCTCGGAACATTTATTTATCAGCACGTTCTCTCTAGCTCCAGACTCTTCGTT 717
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 718 TGGCGCAATTAACAGTAAATACGCTATATGCGCGGTACCTCCATGGCAGACCTATT 777
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 778 GTTCAGGCAATGTCGCCAATTCAGTACGATTTTATAAAAAATAGAGGTATTACTCCT 837
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 838 AAGCCTCTTTAAATAAAGCTCCACTTATCGCTGCTACTGATCTTGGTTAGGATAT 897
Qy 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 898 CCTAATGGTGACCAAGGCTGGGCGGTGTACTCTAAATAAATCGTTAAATGTAGCGTAT 957
Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 958 GTCAATGAAGCAACTGCATTAGCCACAGACAGCAAAAGCAACCTATTCTCCCAAGCACA 1017
Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db 1018 CGCGGTAAACCTTTAAATACTCGTTAGTAGCAGATGCTCTCTGGAAGTACAACGTCA 1077
Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 1078 TCTATACACTAGTATGATTTAGATCTAGTATTACTGCTCCCAATGGCAAAATAT 1137
Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
Db 1138 GTAGGAATGATTTAGTTATCTTATGATAATAAATCGGTGGATGGTCGCAACATGTTGAG 1197
Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrThrIleGluValGlnAlaTyrAsn 420
Db 1198 AAGCTATTATAACGCTCCGCAATCTGGAACGTATATAATAGGTTCAGGCTTCAAGCGTAA 1257
Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1258 GTACCATCTGCGCCACAGCGCTTCTCACTAGCTATCGTACAT 1299

RESULT 9

E03808
LOCUS
DEFINITION
E03808
ACCESSION
E03808.1 GI:2172022
VERSION
JP 1992197182-A/1.
KEYWORDS
Bacillus sp.
SOURCE
Bacillus sp.
ORGANISM
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE
1 (bases 1 to 2218)
Tobe, S., Odera, M. and Asai, Y.
AUTHORS
DNA CODING ALKALINE PROTEASE YA ENZYME AND PRODUCTION OF ALKALINE
TITLE
PROTEASE YA USING THE DNA

JOURNAL Patent: JP 1992197182-A 1 16-JUL-1992;

LION CORP

COMMENT OS Bacillus sp.
 PN JP 1992197182-A/1
 PD 16-JUL-1992
 PF 28-NOV-1990 JP 1990327110
 PI TOBE SEICHI, ODERA MOTOYASU, ASAI YOSHIO
 PC C12N15/57, C11D3/386, C12N9/54, (C12N15/57, C12R1:07), (C12N9/54, C12R1:07);
 PC C12R1:07;
 CC strandedness: Double;
 CC topology: linear;
 CC hypothetical: No;
 CC anti-sense: No;
 CC *source: strain-Y strain;
 CC *source: clone-PUB8A;
 FH Key
 FH Location/Qualifiers
 CDS 218..2125
 /product="precursor of alkaline protease Ya"
 FT sig_peptide 218..2125
 FT mat_peptide 824..2122
 FT /product="precursor of alkaline protease Ya".

FEATURES

Source

1..2218 Location/Qualifiers
 /organism="Bacillus sp."
 /mol_type="genomic DNA"
 /db_xref="taxon:1409"
 707 a 376 c 481 g 654 t
 BASE COUNT 707 a 376 c 481 g 654 t
 ORIGIN

Alignment Scores:

Pred. No.: 1,2e-116 Length: 2218
 Score: 1986.50 Matches: 379
 Percent Similarity: 94.01% Conservative: 29
 Best Local Similarity: 87.33% Mismatches: 25
 Query Match: 88.41% Indels: 1
 Gaps: 1

US-09-985-689A-1 (1-434) x E03808 (1-2218)

QY 1 AsnAspValAlaArgGlyLeuValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 DB 824 AATGATGTAGCAGAGGGATAGTAAAGCTGTTCACAAACAAATACGGATTATAT 883
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 DB 884 GCACAGGTCACACTAGTTCGAGTACGCGACACAGGCTTAGATACAGGTCTGTAACGATAGT 943
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 DB 944 TCTATGCATGAGCATTCGCGGGAAATCACAGCTCTTTACGCGTTAGGAAGAACTAAT 1003
 QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 DB 1004 AATGCGAGTATCCGAAATGGCATGGCACATGATGACAGGTCTGTTACTTGGTAAT --- 1060
 QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 DB 1061 GCTTTAAATAAGAAATGGCTCCCAAGCTAACTTAGCTTCCAATCTATTATGGATAGC 1120
 QY 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 DB 1121 AGCGGAGGATAGTGGCTTACCATCGAACTTAATAGTATTAGTCAAGCTTGGAAAT 1180
 QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140
 DB 1181 GCTGGAGCAGAAATTCATACTAACTCTTGGGAGCCCAAGTAAATGGAGCGTACACTGCT 1240
 QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 DB 1241 AACTCGAGACAAAGTGGATGAGTATGTTGAAATAATGATATGACGGTACTTTTTCAGCT 1300
 QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180

DB 1301 GGTATGAGGTCCTAATTCAGGAACAATAGTGTCCAGGTACAGGAAAAATGCTATT 1360
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 DB 1361 ACGGTGCGCGCAACGGAACATATCGCCAAAGCTTCGGTTTCGATAGCAGATAACCCAAAT 1420
 QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 DB 1421 CATATTGCACAATTTTCATCGAGAGGAGCTACGAGGGATGGACGAATTAAGCCTGACGTA 1480
 QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 DB 1481 ACAGCTCTCGGAACATTTATTTATCAGCAGCTTCTTCTTAGCTCCAGACTCTTCGTTT 1540
 QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetClyGlyThrSerMetAlaThrProIle 260
 DB 1541 TGGGGAATTTATAACAGTAAATACGGGTATATGGCGGTACCTCCATGAGGTATTACTCT 1600
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 DB 1601 GTTGAGGGAATGTCGCGCAATTCAGTGAGCATTTTATAAAAAATAGAGGTATTACTCT 1660
 QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaAlaAspIleGlyLeuGlyTyr 300
 DB 1661 AAGCCTTCTTTAATAAAGCTGCACCTTATCGCTGCTACTGATGTTGGTTTGGATATAT 1720
 QY 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 DB 1721 CCTAGTGGTGACCAAGGCTGGGCGGTGTTACTCTAGATAAATCGTTAAATGACGGTAT 1780
 QY 321 ValAsnGluSerSerSerSerSerSerSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 DB 1781 GTCAATGAAGCAACTGCATTCAGGCACAGGACAAAAAGCAAGCTATTCGTCCAGCACAA 1840
 QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 DB 1841 GCGGGTAAACCTTTAAAAATCTCGTTAGTATGGACAGATGCTCTCGGAAGTACAACTGCA 1900
 QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 DB 1901 TCTATACACTAGTAAATGATTTAGATCTAGTATTACTGCTCCGAATGGACAAAATAT 1960
 QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
 DB 1961 GTAGGAATGATTTTGTAGTATCCCTATGATATAAATACTGGGATGCTCGCAACATGTTGAG 2020
 QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 DB 2021 AACGTATTTATAAACGCTCCGCAATCTGGAAGCTATATAATTTAGGTTCAAGCGTAAAT 2080
 QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 DB 2081 GTACCATCTGCGCCACAGCGTTTCTCACTAGCTATCGTACAT 2122

RESULT 10

AF268611/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

6006 bp DNA linear BCT 30-MAY-2001
 Uncultured marine group II euryarchaeote EBAC37F11, BAC clone
 37F11, partial sequence.
 AF268611
 AF268611.1 GI:9664575
 uncultured marine group II euryarchaeote 37F11
 uncultured marine group II euryarchaeote 37F11
 Archaea; Euryarchaeota; Marine Group II; environmental samples.
 1 (bases 1 to 60006)
 Beja, O., Suzuki, M.T., Koonin, E.V., Aravind, L., Hadd, A.,
 Nguyen, L.P., Villacorta, R., Amjadi, M., Garrigues, C.,
 Jovanovich, S.B., Feldman, R.A. and Delong, E.F.
 Construction and analysis of bacterial artificial chromosome
 libraries from a marine microbial assemblage
 Environ. Microbiol. 2 (5), 516-529 (2000)

```

MEDLINE
REFERENCE
AUTHORS
21128653
11233160
2 (bases 1 to 60006)
Beja, O., Suzuki, M.T., Koonin, E.V., Aravind, L., Hadd, A.,
Nguyen, I.P., Villacorta, R., Amjadi, M., Garrigues, C.,
Jovanovich, S.B., Feldman, R.A. and Delong, E.F.
Direct Submission
Submitted (17-MAY-2000) R & D, Monterey Bay Aquarium Research
Institute, P.O. Box 628, Moss Landing, CA 95039-0628, USA
FEATURES
Location/Qualifiers
1..60006
/organism="uncultured marine group II euryarchaeote 37Fl1"
/mol_type="genomic DNA"
/db_xref="taxon:133822"
/clone="BAC 37Fl1"
/note="environmental sample"
1484..3178
/note="37Fl1#1"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="S-layer protein"
/protein_id="AAF97179.1"
/db_xref="GI:9664576"
/translation="MEPQQTLEIDIFVQSPSAIAGQVNIKIRVSDSGGQGLEFEI
PVRVSPNFEINASQSFVSEGGYPLAWNTGNTMHSVTITILENLPVGNISGMO
SASLSPHEARPIEIIPTTNDKQPLEIELSIVHPOLGMXVESITVIHNSIFASTP
VLGSGSSQAFSLHNSINPSTVTPDATNQDEYRLILEGINFETISSNDALQF
VNIIGILPQVSACQFISDFQELGRAPLTATVTVSCSVNGDLDEQRLTFTFASTNGE
TIPDLQFTIQRNESRVNISVDQWDEPQGLLIRVGVGNGELEVSIETAVQSRQ
SNMVGITASFSTGDLDIATIRTVYEVLDVNCILITVTSSTSSYRIERIVDEGSGQA
PIYKIQNPSIDRESLQATIDCNPFVDVDDPSDNTATAYVEESQSLVTTNNLLWGT
AITFIYGAVALVMORREAGVLIQENIRSKPKPTNNKQTKETEVIEDDSLEIIE
EMELEEPPSVLVEITSETDLSPSRLDTIRQELDPDVEIVDTITSIEERNSKFFD"
3223..3891
/note="37Fl1#2"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="unknown"
/protein_id="AAF97180.1"
/db_xref="GI:9664577"
/translation="MNQNSSTSAFYTLDCGVSNLDETQALILEIRGNHQLWEKLV
NREAFSTRLLKQGLDEIDDRNKITWDDATLWFTCVRLNDRSAPILGENKNRERFGR
FPHHTSGLOYILENLSIFSQSEGYSDLIEHLEYLTERCSEMTGHKELEGKAGL
MILGYITFEVKAIRLSMLGSGMWVSREEPDQGVREAVRHLNALLMAAERRGAGLIH
RQHE"
complement(3892..4797)
/note="37Fl1#3"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="SAICAR synthase"
/protein_id="AAF97181.1"
/db_xref="GI:9664578"
/translation="MGEMLYQGVKQVWSTDDPDLLEFRNTQISVEDQIPLSLPRK
GSLNRTTAHWKLVQEGVCGTHLVNLAPOCLVKVKEVVEKPAIPDAEWVFP
LEELVPHVYLSGNWRFPQGLFPAESLQDAEYGVKLESEVFEVTTKEAFDRNID
REALQISNITDOEFSIYQAAALKVDEIIEERAANKGLIHVDGKKEFALGGRKRVLV
DFIIGLDLWMDADAYSNAGECIELSKFEVRYTHYINSQHAELKLRDSCAIDPIPA
LPESVLEETASYSSMYERLTSGQF"
complement(4846..5508)
/note="37Fl1#4"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="translation initiation factor 6"
/protein_id="AAF97182.1"
/db_xref="GI:9664579"
/translation="MPIDLTIDHSDQIGIHLSASGVNLFHPHRELTPTTIEKLEQVVG
LELAFISGGSNLIGALLAGNTKGMVADIATESIDALTATGDVVMVEGVNTAGNL
LLVNQGGCVASPSIPKDGDLILSEVLGVDIISTIGQDQVGSGLGVNSGVNLLHPDV
APEVLLLEIILGVPPMVGTVSFGSPYVGAGICASNNGALAGSETTGPELNRIEDALG
LI"
5597..5745
/product="5S ribosomal RNA"
5816..8611
/product="23S ribosomal RNA"
complement(8888..10192)
/note="37Fl1#7: contains Zn-ribbon and TPR-like repeats"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="unknown"
/protein_id="AAF97183.1"
/db_xref="GI:9664580"
/translation="MNECAVCGILYLAGSPCAGCSQSLRAOQDSFSDSAAALPTEVPG
LDDAAQAWYDLEIEAPVEQEEAASSLPFGFGESOTNISRPLFCIGSHSPGCIPEL
SNDSGOEEERVOHVPVIESTNLEPVVALETQHDSPVVSASVEIEPVLAPVAPD
LPLEIEPPVPVPGPRLTALPEVEPTVDTDLNPENSIIEFSGSDQITFAEDVIE
VTFSDLEDTVVHVEHDGENYQOMEKDSYSEKSGQPPDLHPAQLAMVDVGTDSASRAL
LESQFTSIGQDWRSAARPOQIVARFPQDSGAMNNYGIALLQVAVHOSHNDPIEVS
NSATOFEAAILSLUREAVTNNSSQPEIYNLSQALLSGREEKALALLDMAGNESQRES
NFTNLRAAILAQLGRYDEAKRLVLPLOAKDLASGNLLKLPAI"
complement(10248..11969)
/note="37Fl1#8"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="permease"
/protein_id="AAF97184.1"
/db_xref="GI:9664581"
/translation="MGCNLFKSRHPAQLIALISLLFVGTMQFTRSSHAAGGDDIDL
TVTIRDFENQAVYVGLGILLFYVALITEVVRHTLAATLGLGLLAIALNHYVSEAL
SLKAVTMDWEITIGLLGMMVMVGIISHTGVFEWFAVQYKSGGETATLWILCSV
TAVLSAFDNVTMLLTPTVIKIAQVLDLKPILILSEVLFNSIGGAATMIGDPNI
MIGSAMSPDAISGNKDNIANLADSGVTNDFITEMAPGILMTVPSEMLIRWLYKDE
FSGRDRDITELERQAKIKGLKVSAGILTLLVLCFFLHPVHLAVSWALSGALV
MLLATNRHELEPELHVETTLFFAGLVLVHLSLOYMGVTDYIGEVYKAIKFPDD
YRLAAAILILWVSATASAFIDNIPYATMIPVLSLAFELNPLNPLMALAFGACL
GGNTLIGASNVVTAGMSEAGYPISEFFRAGFPVMLLSTFVSYFMILVYVVG
EGALTWKIVLVGILSIIITQYSRHAKGSPAEALVDDDFEIVSTIKSVVITGS
KGESEE"
complement(12019..12282)
/note="37Fl1#9: contains KH domain"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="RNA-binding protein"
/protein_id="AAF97186.1"
/db_xref="GI:9664583"
/translation="MSVPNHIIRQAKDTSLPITMRIGKSGLTDSVVVELEGQLSSRL
VKVKVNRGLYPRDELKTVNMAEKTNSVVFAGNVAVFWRN"
complement(12279..12614)
/note="37Fl1#10: contains Zn-finger; conserved protein"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="unknown"
/protein_id="AAF97185.1"
/db_xref="GI:9664582"
/translation="MGRSRLNRDRKHAHWSQTLISLLKNPSOFTPQORSREALH
LVKVSRRHRIPLPSEVREIMCRCTNLLRYGENATIFRNGHKIQTCHSCTSVRRIPY
RHSQKVKVA"
complement(12624..13274)
/note="37Fl1#11"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="adenylate kinase"
/protein_id="AAF97187.1"
/db_xref="GI:9664584"
/translation="MDEGEROGLNVLFGPPGACKGTQOARMMDATGLPQVSTGDM
RAAKACTSVGIEAKSYMKGALVSDSVLIIDIRLHDDDAKNGVMFGFRTVPOA
EALSEIVEVSAVLDAIDPVERIVERICGYSCTGCTGVYHDTFNPIQNGICGGSFTE
KRRADDVEETVLARLEAYHTQTSPLADEYRAAGIFHHVDGDRDIDVIDITDLLALVLC"

```


QY	273	ValLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGly	292	gene
Db	2267	GTAGATCAATAAATAACCAACCACTGGATCATTAATAAAGCAATTAATGATTAATAAT	2326	mRNA
QY	293	Ala-----	302	CDS
Db	2327	GCTCAGTATTAATAGGTACATTTCAATGATTACATCATCAAGTATTACATATCCATCA	2386	
QY	303	GlyAsn-----	311	
Db	2387	ARCCAAGTTTTCAGAAATTTTGCAGGTGCAAGTTTGTTCAGGTGGGTCTATTAGA	2446	
QY	312	LeuAspLysSerLeuAsnValAlaTyrrValAsnGluSerSer-----	325	
Db	2447	ATGAGTAATGGTTACATGTTGTCAATAATAATAATAGTAATAATAATAATAAACAATCA	2506	
QY	325	-----	325	
Db	2507	GATGGTATCACTAAATTTGATGGTATTGGTGGATTAGATTAAAGATTAGTGAACCAAT	2566	
QY	326	-----SerLeuSerThrSerGlnLysAlaThrTyrrSerPheThr-----	338	
Db	2567	CAATGGAAGAAGAAAGTTTATCAACTGGTCAAAATACATCATATGTTTTCACATATAA	2626	
QY	339	-----AlaThrAlaGlyLysProLeuLys-----	350	
Db	2627	CCATCATCATCAAGTAGTAATAGTGGTAATAATATACCAAGAGTTGTTCCACATTAGT	2686	
QY	351	TrpSerAspAlaProAlaSerThrAlaSerValThrLeuValAsnAspLeuAspLeu	370	
Db	2687	TGCACAGATCCACCATCATATGCAGGTGCAAAATTAATTTAGTAATAATTTAGATTGA	2746	
QY	371	Valle-----	378	
Db	2747	ACATGATATTATTATCGTGAATGGTTCAACTATATTTTATTCATCAAGTGTTCA	2806	
QY	379	GlnTyrrValGlyAsnAspPheThrSerProTyrrAsnAspAsnTrpAspGlyArgAsnAsn	398	
Db	2807	TCATTTTAGA-----TTAGCACCACCACTCAAGATACA-----TTGATAAT	2848	
QY	399	ValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrrThrIleGluValGlnAla	418	
Db	2849	GTTGAAGTATAGTTTCATATATCCACGGAACCAATGACTATATAGATTATGTTAGCGGGT	2908	
QY	419	TyrAsnValProValGlyProGlnThrPheSer	429	
Db	2909	ACAAACGTACCAGTGGGACCAAGATTTCTCA	2941	
RESULT 12	AF466309	5772 bp	DNA linear	INV 05-FEB-2002
LOCUS	AF466309			
DEFINITION	Dictyostelium discoideum serine protease/ABC transporter tagD			
ACCESSION	AF466309			
VERSION	AF466309.1	GI:18496823		
KEYWORDS	Dictyostelium discoideum			
SOURCE	Dictyostelium discoideum			
ORGANISM	Dictyostelium discoideum			
REFERENCE	1	(bases 1 to 5772)		
AUTHORS	Anjard,C. and Loomis,W.F.			
JOURNAL	Evolution of the ABC transporters of Dictyostelium			
REFERENCE	2	(bases 1 to 5772)		
AUTHORS	Anjard,C. and Loomis,W.F.			
JOURNAL	Direct Submission			
FEATURES	1	.5772		
source	/organism="Dictyostelium discoideum"			
	/mol_type="genomic DNA"			
/strain="Ax4"				
/db_xref="taxon:44689"				
<243. >5720				
/gene="tagD"				
<243. >5720				
/gene="tagD"				
/product="serine protease/ABC transporter TagD"				
243. 5720				
/gene="tagD"				
/note="similar to TagB and TagC"				
/codon_start=1				
/product="serine protease/ABC transporter TagD"				
/protein_id="AAL74253.1"				
/db_xref="GI:18496824"				
/translation="MKSNTNIRVLVSLVGLILIFILGIRFEFINKNNNDKIGINRKLE				
FSYTKNNNNNNKQDKLKKDKRILLKNEIDTNIKKNIKKNNKNNKEEIPF				
NFISRLKSNDDMETOFTYRKSHVIVQKHNDETRQFKFLNTDVLDEQPYQ				
SHIVNYIPHDSFLVLMNDEQSNLLSSKEWVSWIGPEEPKTHLNYNEKSGILPVYIK				
LSDSTSLIQWENTLNTSYNSKVKLTILNOKKLKSIYVNCNDESSQSGSLVSS				
EKLVIQWISQESSENYIERSEKFOFANRLSPKAIPTGKOTLVNNDRIDIPLRKGOIL				
SIADTGLDGHCHFFSDSNPIPYNSVNLNHRKVITYIGSLHDNEDYVDHGTHVCGGA				
AGAPEDSLAISFSGLATDAKIAFDLASDPSPNPEVPEDEYSQLYPLYNAGARVH				
GDWSLSIQGLGYSDDAGSIDDFLYTHPDFIILRAAGNNFQYSLLSQATAKNVI				
TVGAQTHESYTDALEVSNEFTAKSTLNSLCQSFDDKCYCTYTAQCCTEYTSAR				
LSGCTSYIKNSYASIFSQPELYNENNTCSFSSKGPTRHGRKLPDIAVPGQYITTSR				
SNKATTDQCGDGLPNTNALLSESGTSNATPLATATTLIRQYLVQYVYPTGSIVES				
NKLOPTQKALKALMTNNAQLNGTFLPSTNTNPSNAVFTDFAGANPVQWGSRLRME				
WLYVESSGVKPKSRWVGIGELGDKKASNKWEYSLSGTQNVSYCYFTYKPSSSGNSG				
GIPRTAVLWTDPPSPYSGAKLNNLNTDNTMTSEFIFYSSNGSSSYNGTKGTL				
PLQDINNVGELIYTPINTKISIFRPIIAGTNPIGPQNFVFGENGFEFDWADCS				
MOCNPDPTQCFIENGVSQTCGDDYLMGRCLVQSCNNNYNYSISDKCKFLSYNYI				
VIIVAGTMSLIITVILIKMEYKENGKFSLEKFEFSGVLGTVKNGYGGKGGSGGS				
SGSGTLKDGTDIDGTHIVREPKDAPVTPPDLYSLSPFIIETLSTACSLVATAA				
SILQPYITQIIOIDPTIGDRDQFIIFLLALLEFVFSTISSWISGLVNEKMYM				
RLQKVFALLAQDFFQKNSAAVLNVLVTPMLRSLTGLLSVGVGICFVGS				
LVFTISWKLSLAFATVLAIVTQVQSKFTRTLRLLFHNKSAQSHQESVMVN				
HVNSVCQDREIAKYSEOLMVVQISRLIINNTFAASIKWLAVESLAFILIFYGAY				
LAIQKQTVGLVLSPLYICYVIDSTTLFGVYSYVOCIASATRVFLILRSAPRKRT				
TLQKQFVGLVLSPLYICYVIDSTTLFGVYSYVOCIASATRVFLILRSAPRKRT				
LEDEELNIDTNOQNNNNNDJISDSSDDDDNNNNNNNNNNNNNNNNNNNNNNNN				
EDAEYKNNKNNNNKMTTLKSNPPLVGVGIDNNNNNNNNNNNNNNNNNNNNNNNN				
EIDDDGDDDDDDDEGDDNNNNNNNDPNDNGIEMLETKQLKRRKQMKKEFYKKT				
GISLELNLIPISAYTITECEKIEFKYSCYPYRADVGLVNLIDLFESGKCYGLV				
GPSGSGKSTLLELISFYSGHSGGKIYMDGIDIAKIRPSNLRSFVNLVHOHFLFA				
TIENIGVALDNPOTORDITEAALANAHEFIOSLPKOYDTMLTDGGLSGGQKKRIAV				
ARAIKAKRIMLLDEITAELEPSEEAINKIKVLTGHTVMVAHVAAVARDCKIF				
VLDKQGVIEQGHQNLQMLKMKKYRFPASEDDDDYAPLLVL"				
BASE COUNT	2189 a	708 c	904 g	1971 t
ORIGIN				
Alignment Scores:				
Pred. No.:	6	51e-23	Length:	5772
Score:	511.50	Matches:	164	
Percent Similarity:	40.82%	Conservative:	76	
Best Local Similarity:	27.89%	Mismatches:	155	
Query Match:	22.76%	Indels:	193	
DB:	3	Gaps:	22	
US-09-985-689A-1 (1-434) x AF466309 (1-5772)				
QY	19	LeuTyrrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg	37	
Db	1221	TTGAGAGGTAAGGTCAGATATTGAGTATTCGCCATCTAGTGGTGGTATGATGTT	1280	
QY	38	-----AsnAspSer-----	48	
Db	1281	TTCTTTTCAGATTCATAATAATCCATATACCATATAATAGTGTAAATTTAAATCATAGAAAA	1340	
QY	49	LysIleThrAlaLeuTyrrAlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsnGlyHis	68	
Db	1341	GTGTAACTTATATGTTCTCTA-----CATGATAATGAGGATTATGTCGATGCTCAC	1394	
QY	69	GlyThrHisValAlaGlySerValLeuGlyAsnGly	81	
Db	1341	GTGTAACTTATATGTTCTCTA-----CATGATAATGAGGATTATGTCGATGCTCAC	1394	

US-09-985-689A-1 (1-434) x DDU20432 (1-6115)

19 LeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg--- 37
|||||
|||||

204 TTATTAAGCTTAAACCCGTCGATATTGCTACTGGTTTAGCATCGTGACCAATTGT 1353

```

QY 299 GlyTyr----- 300
Db 2413 ACTTATTTTGGAGTCATCAAGTACAAATCCATCAACGCAATATTGACAAATTAAT 2472
QY 301 ---ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAla 319
Db 2473 GGCACAATTAATTAAGGTTGGGTGCGTCACTAGATGAAT-----AATTGGTTA 2523
QY 320 TyrValAsnGluSerSer-----SerLeuSer 328
Db 2524 TATGTTAAATCATCAATATCCACACCACCATCAAGATGGATTGGTATCGGTTAGGT 2583
QY 329 ThrSerGlnLysAlaThr----- 334
Db 2584 AAAAATCAAAAGCAACTGAATGAAGAAAGATAGTTTATCAAGTGGTTAAATAAATCA 2643
QY 335 TyrSerPheThr-----AlaThrAlaGlyLys 343
Db 2644 TATTGTTTCACATATAAACCATCATCATCATCATCATCATCATCATCATCATCAT 2703
QY 344 Pro---LeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerVal 362
Db 2704 CCAAGAATTGTCGAACATTAGTTGGACAGATCCCGTCATATTCTGGTGCAAAATT 2763
QY 363 ThrLeuValAsnAspLeuLeu-----ValIleThr 373
Db 2764 AATTTAGTAATAATTTAGATTATTATTATTAAATAGTATGATGATTCAATAATATCA 2823
QY 374 AlaProAsn-----GlyThrGlnTyrValGlyAsnAspPheThrSerProTyrAsnAsp 391
Db 2824 AUTGGTAATCTGGTGGATCATTAACACACGACGAGGAAGTAGCACAACTT----- 2874
QY 392 AsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThr 411
Db 2875 -----GATACATTGAATAATGTTGAGGTATATAATTAATCAACCAAGCAATGAAT 2928
QY 412 ThrThrIleGluValGlnAlaTyrAsnValProValClyProGlnThrPheSer 429
Db 2929 TATAAATTTACATCCGACGAACCAATGTACCAATGGACCAAAAGTTTTCAT 2982

RESULT 14
AR201152
LOCUS AR201152 1977 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 11 from patent US 6358726.
ACCESSION AR201152
VERSION AR201152.1 GI:20252040
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1977)
AUTHORS Takakura,H., Morishita,M., Shimojo,T., Asada,K. and Kato,I.
TITLE Thermostable protease
JOURNAL Patent: US 6358726-A 11 19-MAR-2002;
FEATURES
    Location/Qualifiers
        source
            1..1977
                /organism="unknown"
BASE COUNT 453 a 659 c 554 g 311 t
ORIGIN

Alignment Scores:
Pred. No.: 9.25e-20 Length: 1977
Score: 452.50 Matches: 138
Percent Similarity: 44.66% Conservative: 67
Best Local Similarity: 30.07% Mismatches: 153
Query Match: 20.14% Indels: 101
DB: 6 Gaps: 18

US-09-985-689A-1 (1-434) x AR201152 (1-1977)

QY 8 ValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnIleValAla 27
Db 433 ATAGGGCCGATACCGCTGTGGAACCTCCCTCGGCTACGACGAAGCGGTGTGGTGTGCC 492

```

```

QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisLluAlaPheArg 47
Db 493 ATCGTGCATACGGGTATAGACCGCAAC-----CACCCGATCTGAAG 534
QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn 66
Db 535 GGCAGAGTCAATAGGCTGGTACGACGCGCTCAACGGCAGGTGCGACCCCTACGATGACCCAG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLys----- 84
Db 595 GGACACGGAAACCCACGTTGCGGGTATCGTTCGCCGAACCGCGCATCGTTAACTCCACGTAC 654
QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 101
Db 655 ATAGGCTGCGCCCGCGGCAAGCTCGTCGGGTCAAGGTTCCTCGGTGCGCGCGGTTCG 714
QY 102 GlyGlyLeuGlyClyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 121
Db 715 GGAAGCGTCTCCACCATCATCGCGGTGTTGACTGGTCTGCCAGAACAGCAAGTAC 774
QY 122 GlyAlaArgIle-----HisThrAsnSer 129
Db 775 GGGATAAGGTCATCAACCTCTCCCTCGGTCTCCAGAGCTCCGAGCAAGCAAGTAC 834
QY 130 TrpGlyAlaAlaValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrVal 149
Db 835 CTCAGTCAGCGCGTCAACACGCGCTGGGAGGCC----- 867
QY 150 ArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThr 169
Db 868 -----GGTATAGTCTGCGTCGCGCGCGCAACAGCGGCGCAACACTACAC 918
QY 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg 189
Db 919 CTCGCTCACCCCGCGCGCAAGGTCATAACCGTCGTCGCA----- 963
QY 190 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 209
Db 964 -----GTTGACAGCAACGACATCGCGAGCTCTCCACAGCGGA 1005
QY 210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSer 229
Db 1006 CCGACCGCGGACGGAAGCTCAAGCGGAGTCTGCGCGCGCTGACATCATAGCC 1065
QY 230 AlaArgSerSerLeuAlaProAspSerPheTrpAlaAsnHisAspSerLysTyrAla 249
Db 1066 CCGCGCGCGCAGC-----GGAACGACATGGGCACCCCGATAACGACTACTACAC 1116
QY 250 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly---AsnValAlaGlnLeu 268
Db 1117 AAGGCTCTGCAACACGATGGCACCCCGACGTTTCGGGCGTGGCGGCTCATCTCT 1176
QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLys 286
Db 1177 CAGCCCGAC-----CCGAGTGCACCCCGCAAGGTGAAG 1212
QY 287 AlaAlaLeuIleAlaGlyAla-----AlaAspIleGlyLeu 298
Db 1213 ACCGCTCTCGACACCGCGCATAGTCGCCCAAGGACATAGCGGACATCGCTAC 1272
QY 299 GlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeu----- 316
Db 1273 GGTGCG-----GGTAGGTGAACGCTACAGGCCATCAAGTAC 1311
QY 317 ---AsnValAlaTyrValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyr 335
Db 1312 GACGACTACGCCAAGCTCACTTCACGGCTCGCTCGCGCAAGGAAGGCCACCCAC 1371
QY 336 SerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaPro 355
Db 1372 ACCTTCGACCTACGCGCGCACCTTCGTGACCGCCACCTCTACTGGGAC----- 1422

```

FEATURES
SOURCE

gene
CDS

CDS

```

/gene="taga"
/note="putative serine protease/ABC transporter"
/codon_start=1
/protein_id="taga"
/db_xref="GI:9992887"
/translation="MNKKLPIFGLSFLFIFNLSLSLKSKNFYNHNHINPIQIH
RNEGNKLSLKLHNDVITISNRONILFNKLSNQKSGSLFLNHLNPIEQV
HNNEGILQDLFNGEITFIYPTDNTITLSTGSDNNNNNNKILINKEKLPIPSIQ
WKLPELRKILPSLFQNGQDNQNEIDQRIYIYHENSNOQSDINDITISESSLFLV
EKELISNNNNNNNVLTVNLSKQSKLSIEIYIKTSLRVLVYSPSSSKLIKHTESP
KFAHYISQSGASPTSPIDVJIGKGIEVGCADTIGIDINCHCFDPISTGTHRK
LISYSSNGQDIDRIGGTHIVGTILGSTVYSEFSGGAKSVKFAVDLVQSGV
NGLISQNLTAIYOSTQDNKAKVCDANNLSIGPYGTETMDRQWDHPDLFVLR
AGGNYVFGTISYTLTSGESTSKSLVSGSQNPSTYSSDIYDWDWDFYNSRTSVC
TOSGVNIGTISDYPTTQDITQCSNPILATKICITSEIOQQYQNTSVTSEFIP
SLFSGVGTSDGRULPDLLAPGSITISRSRGLSPSTINHCSTIGTISALIAMGSS
SQAAVATSAALVRYQRYDGFYINGKYSVSGFOPASLVRKATLITASINDTLE
YSCFGDNIQKSLTITNAQTLSDTPSIEKADPIINTGTNSYCFSLDSKADIDIL
LVOTFAGPSPLSTFLVNNLDLALAFDEKELSIYSGNETSIFKNTSOFIDQNNVE
VIRIKDAPISGYDVKIEGTNIIVPNQSYSVLTSGGTTLMKESECAOCFYDNDQGS
OCMEFDGIGTGYCKDDNRFSKVCYEGNTGYVFDNGITKSCVTTLALYDVLVLLGI
FGIILVAGFVLVYKSKSLDKNYKIFSLSKDQGDGNSIRSVSAGNNNNNNNNNN
PQYDNNNNNNNNNNNNNNNNNNNSKQSIENLSVGGDDGTGPDGDDQQQNNNS
POYDQDGRILSQGEVEISIFEYISLQPEKSTIGLALFSLFDVAGLVLVAAANFI
DYLAVGETKGTSTLTIPALIIIGMIVFGSLLIALAGHRIAREMFASLKLQKQ
OMAFENRKTGELMSRIASDVSYSVRSIIVSDPSIMHIIQIATIGTITLFIISWLKLS

```

BASE COUNT 1843 a 716 c 792 g 1758 t

ORIGIN

VVLCPILLVFSKPYGVIEISVKVQDALADAAATHAAETLFWNKTVWVFSAREERE
AKFKSLISGVYKALMTWNGTYSKSGFSEQLSVFILLWYSSGVNGDUTSPMLI
AFNLFPIITGAVTQVASYLTYSKSGSYRFEIMQRPVDIOGEGITRTKVRGDI
QFNKVSFAYSNNPDOLVLEKIDIKPEPPTITALIGPSGGKSTMLSLCRLYNIDGGS
ITLDGTNKFDPVNLHRLHSIVNPQPSLFSGSIAENIMCKPTATRSEI IEACKQAN
AHDEITAMPEGYDTLIGERSTALSGGQKRIARTI IKNPTVLLDDETTSELDVESE
KLVQSDIKLVGWRTVYIVAHRLTITLADIIVVSDSTISEMTQPEELLAKKGMFYD
FVQIYQGGEBELDIQPSNRNTRNADKLNRNSETIKQIAKINNI IPIHRPQRGDD
NEDDENNSGGSSRSPHOCQCDKPRRQMLNQCY"

Alignment Scores:

Alignment Scores:	
pred. NO.:	6.68e-19
Score:	447.00
Percent Similarity:	44.01%
Best Local Similarity:	25.78%
Query Match:	19.89%
DB:	3
Length:	5109
Matches:	140
Conservative:	99
Mismatches:	164
Indels:	140
Gaps:	22

US-09-985-689A-1 (1-434) X AF263455 (1-5109)

QY 18 GlyLeutyGlyClnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
 ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
D6 898 GGAATTAAAGGTGATCGTGAATTTGGTTTCAGATACTGGTATTGCATTAATCAT 957

[illegible]

958 TGTTCCTTTAAGATACAAATCCAAATGGTTCGACCTAAGAAATCATTCCTA
Db

Qy 55 AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGly 74

Db 1015 TCAAGTGGAAATGGTGATCAAAATCGATGAAATTGATGGTCATGGTACTCATATTGTTGGT 1074

QY 75 SerValLeuGlyAsnGlySerThrAsn-----LysGlyMetAlaPro 88

1075 ACTATTATAGGCTCAACTACTGTGGATCCAAGTGTTTCAGAAATTAGTGTGGCGCTCCA 1134

89 ClnAlaAsnIeuValpheGlnSerIle---MetAspSerGlyGlyLeuGlyGlyLeu 107

137

[illegible]

DB 119Z CAATCAAAATTATACAGCATACTCATTCAGCCGTTTGTGTAATCGGGGTTTTTTT

Qy 128 AsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrTrnAspSerArgAsnValAspAsp 147

Db 1252 GATGATGGAATTCAAATATTTGGTCCATTCTATCTGCTCACTGAAATGATAGATAGA 1311

QY 148 TyrValArgLysAsn---AspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsn 166

db
1312 TTCCAATGGGACCACCCAGATTTCCTTGTTAGTTAGATCAGCAGGAAAT--AATGTAAC 1368

167 C L V C L V ----- Thr T L e Ser Ala pro Gly Thr Ala Ala Lys Asn Ala T I e Thr 181

[illegible]

1000

[illegible]

DB
1429 GTTGGT⁺AGT⁺C⁺AAAT⁺-----CAACCATCA⁺-----AGTACCTATTAATCATGCTGCTGCTTAA⁺

QY	201	201

Db 1480 TGGGACTGGGATTTCATTTATTAATTCAATTAGAACTTCAGTTTGTACTCAAGGTCAATCT 153

QY 201 201 201

1540 ATTTATGGTATAACTTGTTCAGATGTACCAACTCAAACAACCTCTGTCGATATTCAAACCT 159

201 201

[illegible]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

```
QY 202 -----ValAlaGlnPheSerSerArgGly 209
Db 1650 TATCAACCAATTCAACAGTTTATAGTGAATTCATCCAAAGTTTATTTAGTGGTGTGGT 1719
QY 210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSer 229
Db 1720 CCAACATCTGATGTCGTTTGAACACAGATTATTAGCACCAGGTTCCACCAATTATTTC 1779
QY 230 AlaArgSerSerLeuAlaProAspSerPheTrpAlaAsnHis----- 244
Db 1780 TCACGT---TCCTTGGGTCCATCATCAACCAATT-----AATCATGTGTCACCAATTACA 1830
QY 245 -----AspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 1831 TCAGGTATTGCAACATCGCTTTAATCGCTATGGAAGGTTCAAGTCAAGCAGCTGCTGTC 1890
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGlu-----HisPheValLysAsn 275
Db 1891 GCTACAAGTCCCGCAGTTTGTGTCGTCATTAATCTACTATAGAGATGGTTATTTTATTAATGGT 1950
QY 276 Arg-----GlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIle 290
Db 1951 AAAGTTAATTCATCTGTGGGATCCCAACCATCAGCATCATTTGGTTAAAGCAACCTTAATC 2010
QY 291 AlaGlyAlaAla-----AspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrp 307
Db 2011 AATACTGCAAGTATTAAATGTGTGATTCACATTAAGATAT-----TCACAAGGATTT 2061
QY 308 GlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSer----- 325
Db 2062 GGTAATATTCAACTTTCAAAGTTAATCACAACACTACAATGCCAAACCAACCATCTTTAGAC 2121
QY 326 -----SerLeuSerThrSerGlnLysAlaThrTyr 335
Db 2122 ATTCATCAATCAATGAAAAAGCTGATCCAAATAATTAATACTGCTGGAACCAATTCATAT 2181
QY 336 SerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaPro 355
Db 2182 TGTTTCCTCACTAGTAAGCAGATATTGATATAACATTTGGTTGGACAGATCCTGCC 2241
QY 356 AlaSerThrThrAlaSerValThrLeuValAsnAspLeuAspLeuValIleThrAla--- 374
Db 2242 GGTTCACCATTAACAACATTTACTTTGGTAAATAATTTGGATTAGCTTTACTTGCATTT 2301
QY 375 ProAsnGly-----ThrGlnTyrValGlyAsnAspPheThrSerProTyrAsnAspAsn 392
Db 2302 GTTGATGGTGCACTTTCAATATATTCAGGAATAGTGAACCACTTTTCAAAAAATACAAGC 2361
QY 393 -----TrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSer 409
Db 2362 CAAGTAATTTTCGATCAATTAATAATGTTGAAGTCATTAGAATTAAGGATGCACCAATT 2421
QY 410 GlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyProGlnThrPheSer 429
Db 2422 GGATGATTATGATGTTAAATAATTTGTTACAAATATTTGTAATTCCAAATCAATCCTATTCA 2481
QY 430 LeuAlaIle 432
Db 2482 GTTGTAAATT 2490
```

Search completed: July 28, 2003, 07:02:57
Job time : 5474.74 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 28, 2003, 03:42:25 ; Search time 100.115 Seconds
(without alignments)
1913.396 Million cell updates/sec

Title: US-09-985-689A-1

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSYGLY.....EQAYNPVGPQTFSLAIVN 434

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model
-Q=/cgn2_1/USPTO_spool/US09985689/runat_25072003_143032_864/app_query.fasta_1.1166
-DB=Issued Patents_NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09985689 -CGN_1_115=@runat_25072003_143032_864 -NCPU=6 -ICPU=3
-NO_MMAPP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents_NA:*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*\n2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*\n3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*\n4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*\n5: /cgn2_6/ptodata/1/ina/PTCUS_COMB.seq:*\n6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2247	100.0	1923	4	US-09-509-814A-5
2	2242	99.8	1923	4	US-09-509-814A-7
3	2183	97.2	1920	4	US-09-509-814A-3
4	2125.5	94.6	3003	2	US-08-873-479-41
5	452.5	20.1	1977	3	US-08-894-818B-2
6	452.5	20.1	1977	4	US-09-445-472-11
7	416.5	18.5	1236	3	US-09-445-472-2
8	416.5	18.5	1566	3	US-08-894-818B-4
9	416.5	18.5	1962	3	US-08-894-818B-34
10	416.5	18.5	1962	4	US-09-445-472-15
11	403.5	18.0	1977	3	US-08-894-818B-6
12	351	15.6	2539	3	US-09-000-016-3

13	351	15.6	2539	4	US-09-514-340-3	Sequence 3, Appli
14	346	15.4	2809	3	US-09-000-016-1	Sequence 1, Appli
15	346	15.4	2809	4	US-09-514-340-1	Sequence 1, Appli
16	310.5	13.8	2532	1	US-07-671-376C-4	Sequence 4, Appli
17	307.5	13.7	1859	3	US-08-894-818B-15	Sequence 15, Appli
18	307	13.7	2835	1	US-08-750-532-2	Sequence 2, Appli
19	307	13.7	4765	1	US-08-750-532-8	Sequence 8, Appli
20	307	13.7	4765	3	US-08-894-818B-7	Sequence 7, Appli
21	307	13.7	4765	4	US-09-445-472-5	Sequence 5, Appli
22	296	13.2	898	1	US-08-750-532-7	Sequence 7, Appli
23	290.5	12.9	564	1	US-08-750-532-11	Sequence 11, Appli
24	290.5	12.9	564	3	US-08-894-818B-14	Sequence 14, Appli
25	276	12.3	1306	4	US-09-966-921A-1	Sequence 1, Appli
26	276	12.3	1330	4	US-09-966-921A-5	Sequence 5, Appli
27	256.5	11.4	1194	4	US-09-328-352-3407	Sequence 3407, Ap
28	256.5	11.4	10216	2	US-08-875-154-1	Sequence 1, Appli
29	252	11.2	1628	4	US-09-634-238-212	Sequence 212, App
30	251	11.2	840	1	US-08-434-255-7	Sequence 7, Appli
31	251	11.2	840	1	US-08-459-967-7	Sequence 7, Appli
32	251	11.2	840	1	US-08-460-327-7	Sequence 7, Appli
33	251	11.2	840	1	US-08-459-871-7	Sequence 7, Appli
34	251	11.2	840	3	US-09-024-532-1	Sequence 1, Appli
35	251	11.2	840	4	US-09-104-623A-1	Sequence 1, Appli
36	251	11.2	840	4	US-09-019-532-1	Sequence 1, Appli
37	251	11.2	840	4	US-09-417-359A-1	Sequence 1, Appli
38	251	11.2	1110	1	US-08-434-255-5	Sequence 5, Appli
39	251	11.2	1110	1	US-08-459-967-5	Sequence 5, Appli
40	251	11.2	1110	1	US-08-460-327-5	Sequence 5, Appli
41	251	11.2	1110	1	US-08-459-871-5	Sequence 5, Appli
42	251	11.2	1191	1	US-08-434-255-3	Sequence 3, Appli
43	251	11.2	1191	1	US-08-459-967-3	Sequence 3, Appli
44	251	11.2	1191	1	US-08-460-327-3	Sequence 3, Appli
45	251	11.2	1191	1	US-08-459-871-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-09-509-814A-5
; Sequence 5, Application US/09509814A
; Patent No. 6376227

; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO

; APPLICANT: OKUDA, MITSUYOSHI

; APPLICANT: SAEKI, KATSUHIKA

; APPLICANT: KUBOTA, HIROMI

; APPLICANT: HITOMI, JUN

; APPLICANT: KAGEYAMA, YASUSHI

; APPLICANT: SHIKATA, SHITSUW

; APPLICANT: NOMURA, MASAFUMI

; TITLE OF INVENTION: ALKALINE PROTEASE

; FILE REFERENCE: 0327-0832-0PCT

; CURRENT APPLICATION NUMBER: US/09/509,814A

; CURRENT FILING DATE: 2000-04-06

; PRIOR APPLICATION NUMBER: PCT/Jp98/04528

; PRIOR FILING DATE: 1998-10-07

; PRIOR APPLICATION NUMBER: JP 9-274570

; PRIOR FILING DATE: 1997-06-08

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: Patentin version 3.0

; SEQ ID NO 5

; LENGTH: 1923

; TYPE: DNA

; ORGANISM: Bacillus sp.

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(1923)

US-09-509-814A-5

Alignment Scores:

Pred. No.: 1.94e-221

Score: 2247.00

Percent Similarity: 100.00%

Length:

Matches: 1923

Conservative: 434

Handwritten: Nucleic acid

```
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-09-985-689A-1 (1-434) x US-09-509-814A-5 (1-1923)

QY 1 AsnAspValAlaArgGlyLeuValAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
DB 619 AATGATGTTGGCGTGGAAATGTTCAAGCGGATGGCTCAGAGCAGCTACCGGTTGTAT 678
QY 21 GlyGlnGlyGlnLeuValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 679 GGACAGGACAGATCGTAGCGGTTGCCGATACAGGCTGTGATACAGTGCAGTACAGT 738
QY 41 SerMetHisGluAlaPheArgGlyLysLeuThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
DB 739 TCGATGATGAAGCTTCGGCGGAAATTAATGCTGATATATGCAATGGGACGCGAGAT 798
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
DB 799 AATGCCAATGATACGAATGGTCTATGTCGCGCATGGCTGGCTCCGATATTAGGAACGGC 858
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
DB 859 TCCACTAATAAGGAATGGCGCTCAGGCGAATCTAGTCTTCCAATCTATCATGGATAGC 918
QY 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
DB 919 GGTGGGGACATGGAGGACTACCTCGAATCTGCAAACTTATTCAGCCAGGATACAGT 978
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAlaAsnGlyAlaTyrThr 140
DB 979 GCTGGTGCAGAAATTCATACAACTCTCTGGGAGCAGCAGTGAATGGGGCTTACACAACA 1038
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
DB 1039 GATTCAGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCCTTTTCGGTGGC 1098
QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
DB 1099 GGAATGAAGGACCGACCGGGAACCATCAGTGCACCGCAGCAGCTAATAATGCAATA 1158
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
DB 1159 ACAGTCGGAGCTACGGAACCTCCGCCCAAGCTTTTGGTCTTATGGCGCAATATCAAC 1218
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
DB 1219 CATGTGGCACAGTCTCTTACGTGGCCGACCAAGATGGAGGATCAACACCGATGTC 1278
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
DB 1279 ATGGCACCGGAACGTTTACTATATCAGCAAGATCTTCTCTGACCGGATTCCTCTC 1338
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
DB 1339 TGGGGCAACCATGACAGTAAATATGATATGATGGTGGAACTCATGGCTACACCGATC 1398
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
DB 1399 GTTGCTGGAAACGTGGCACAGCTTCGTGAGCATTTTGTGAAACACAGAGGATCACACCA 1458
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
DB 1459 AAGCCTTCTCTATTAAGCGGACCTCATGTCGCGTGCAGCTGATCGGCTTGGCTAC 1518
QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
DB 1519 CCGAACGGTAACCAAGGATGGGACGAGTGACATTTGGATAAATCCTCGAAGCTTCCCTAT 1578
QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
DB 1579 GTGACAGGCTCCAGTCTCTATCCACCAGGCCAAAGGACGCTACTCGTTTACTGCTACT 1638

RESULT 2
US-09-509-814A-7
; Sequence 7, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, MATSUHISA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUM
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 1923
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1923)
US-09-509-814A-7

Alignment Scores:
Pred. No.: 6,35e-221 Length: 1923
Score: 2242.00 Matches: 433
Percent Similarity: 99.77% Conservative: 0
Best Local Similarity: 99.77% Mismatches: 1
Query Match: 99.78% Indels: 0
DB: 4 Gaps: 0

US-09-985-689A-1 (1-434) x US-09-509-814A-7 (1-1923)

QY 1 AsnAspValAlaArgGlyLeuValAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
DB 619 AATGATGTTGGCGTGGAAATGTTCAAGCGGATGGCTCAGAGCAGCTACCGGTTGTAT 678
QY 21 GlyGlnGlyGlnLeuValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 679 GGACAGGACAGATCGTAGCGGTTGCCGATACAGGCTGTGATACAGTGCAGTACAGT 738
QY 41 SerMetHisGluAlaPheArgGlyLysLeuThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
DB 739 TCGATGATGAAGCTTCGGCGGAAATTAATGCTGATATATGCAATGGGACGCGAGAT 798
```


Tue Jul 29 14:12:39 2003

us-09-985-689a-1.p2n.rni

ADDRESSEE: No. 5891701o No. 5891701disk of No. 5891701th America
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/873,479
FILING DATE: 12-JUN-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Agilis, Cheryl H
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 5251.000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:

INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 3003 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-873-479-41

Alignment Scores:
Pred. No.: 1,24e-208 Length: 3003
Score: 2125.50 Matches: 406
Percent Similarity: 97.93% Conservative: 19
Best Local Similarity: 93.55% Mismatches: 8
Query Match: 94.59% Indels: 1
DB: 2 Gaps: 1

US-09-985-689a-1 (1-434) x US-08-873-479-41 (1-3003)

976 GCAGGTGCCAGATTATCATCAAACTCTCTGGGGGCGAGCGGTGAATGGGCGCTACACGACA 1035
141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
1036 GATTCCAGAAATGGATGACTATGATAGAGAAATGATATGACGATTTCTTTTCGCGGCT 1095
161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
1096 GGGATGAAGCCGACGCGGTACCATCAGTCACCTGGTACGCTAAACAGCCATA 1155
181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
1156 ACAGTCGCGCAACCAAACTCGCTCAAGCTTCGGTTCCTATGAGATATATATAC 1215
201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
1216 CACGTGACAGTCTCTCCGTCGCGCGACAAAGATGGGCAATCAAGCCGTGATGTC 1275
221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
1276 ATGGCGCCAGGACATACATTTATCAGCAAGATCTCTCTGCACCCGATTCCTCTTC 1335
241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetClyGlyThrSerMetAlaThrProIle 260
1336 TGGCGAATCATGACAGCAAAATATGCTATATGGTGGACGCTCCATGGCAACACCGATT 1395
261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
1396 GTTGGCGGAATGTTGCACAGCTCCGTGAGCATTTGTGAAAAATAGAGGAATCACCTCT 1455
281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
1456 AAGCCTTCCTATTTGAAGCAGCTTTGATTCAGGCTGCTGATGTGGATTGGGTTAT 1515
301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
1516 CCGAAGCAAAACCAAGGATGGGCGCGAGTACACCTGGATAAATCGTTGAACGTTGCCAT 1575
321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
1576 GTGAACGAATCCAGTCCCTATCAACTAGCCAAAGGACATATACCTTTACTGCAAG 1635
341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
1636 GCGGGCAAGCCATGAAATCTCCCTGATGTGCGGATGCTCCCTGCAAGCACTACTGCT 1695
361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
1696 TCTGTAACTGCTCAATGATTTGGATTTGGTCTATTACAGCACCACCAAGCAAGATAT 1755
381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
1756 GTCGGGAATGACTTCTCAGCACCATTTGACAAATTAACCTGGATGGCGCGCAATAC 1815
401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
1816 AATGTATTATTAAATTCGCCCAAGTGAACATATACCATTTGAGGTGCAAGCATATAAT 1875
421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
1876 GTGCGGTTGGACCAAACTTCCTGTTGGCAATTTGGAAC 1917

RESULT 4

US-08-873-479-41
; Sequence 41, Application US/08873479
; Patent No. 5891701
; GENERAL INFORMATION:
; APPLICANT: Sloma, Alan
; APPLICANT: Lynne, Christanson
; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide
; TITLE OF INVENTION: Having Protease Activity
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:

```

Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1947 GGAATGAGGACAGGAGTAGCGGTACATCAGTCAGCAGGAGCAACAAATGCGATT 2006
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 2007 ACAGTTGGGCAACGAAACCTAGTCCAGCTTCGGATCTTATGCGGATATATTAC 2066
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 2067 CATGTTGCTCAATTCCTCAGGAGGCTCTACTAGATGGACGTATTAAAGCGGACGTC 2126
Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
Db 2127 ATGCAACAGGTACGTATATCTCTGCTAGATCATCATTTAGTCCAGATTCCTCATTC 2186
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 2187 TGGCAACCATGATAGTAATATGCTACATATGCTAGTGGTGTCTTCTATGGCTACTCCAAT 2246
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 2247 GTAGCAGGTAATGTGCACATTAAGGAGCATTTTGIGAAAAATAGAGGGGTAACTCCT 2306
Qy 281 LysProSerLeuLysAlaLeuLeuAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 2307 AAGCCTTCCCTTTTAAAGCTGCTTTAATGTCAGGTCTGGGATGTTGGACTTGGCTTT 2366
Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 2367 CCAATGGTAACCAAGATGGGAAGATACGTTAGATAAATCCCTAAATGTGCGATTT 2426
Qy 321 ValAsnGluSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 2427 GTGAATGAAAGAGCCCTTTTATCAACAAGTCAAAAACACATATTCGTTACGGCTCAA 2486
Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db 2487 GCTGGTAACCCCTTAAAAATATACATCTGTTGTCAGATCAGCAGGTAGCAGCGCA 2546
Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 2547 TCACATAACTTGTAGTATGATTTAGATCTAGTAAATCAGTCACCAAAATGGAACCTAAATC 2606
Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
Db 2607 GTCGAAATGACTTTACAGCAGCGTATGATAACAATGGGATGGCAGAAACACGTGGA 2666
Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 2667 AATGTTTATCAATGCTCCTCAAGCGGAACGTATACAGTCGAAGTGCAGGCTTACAAT 2726
Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 2727 GTACCAGTAAGTCCGCAACCTTTCTTTTACGATTTGTACAT 2768

```

RESULT 5

```

; US-08-894-818B-2
; Sequence 2, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:

```

```

; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPER-THERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark

```

```

; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-894-818B-2

```

Alignment Scores:

```

Pred. No.: 9,03e-37 Length: 1977
Score: 452.50 Matches: 138
Percent Similarity: 44.66% Conservative: 67
Best Local Similarity: 30.07% Mismatches: 153
Query Match: 20.14% Indels: 101
DB: 3 Gaps: 18

```

```

US-09-985-689A-1 (1-434) x US-08-894-818B-2 (1-1977)

```

```

Qy 8 ValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnIleValAla 27
Db 433 ATAGGGGCGGATACCGTCTGGAACCTCCCTCGGCTAGCAGCGAAGCGGTGGTGGTGGC 492
Qy 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
Db 493 ATCGTCGATACGGGTATAGACGCGAAC-----CACCCCGATCTGAAG 534
Qy 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn 66
Db 535 GGCAAGGTCATAGGTGGTGTGACGCGCTCAACGCGAGGTGACCCCTACGATGACAG 594
Qy 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLys----- 84
Db 595 GGACAGGAACCCCGTTCGGGTATCGTTCGGGTCAGGTTCCTGGTGGCCACGCTTGG 714
Qy 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 101
Db 655 ATAGCGGTGCCCCCGCGGCGAAGCTCGTCGGGTCACAGGTTCCTGGTGGCCACGCTTGG 714
Qy 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 121
Db 715 GGAAGCGTCTCCACCATCATCGCGGGTGTGACTGGGTCTGCCAACAGACAGCATAC 774
Qy 122 GlyAlaArgIle-----HisThrAsnSer 129
Db 775 GGGATAAGGGTCAATCAACCTCTCCCTCGGCTCTCCAGAGCTCGGAGGAACGACTCC 834

```

APPLICANT: ASADA, Kiyozo
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
FILE REFERENCE: TAKAKURA-6
CURRENT APPLICATION NUMBER: US/09/445,472
CURRENT FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 151969/1997
PRIOR FILING DATE: 1997-06-10
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn version 3.0
SEQ ID NO 11
LENGTH: 1977
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic
US-09-445-472-11

130 TrpGlyAlaAlaValAsnGlyAlaTyrThrAspSerArgAsnValAspTyrVal 149
835 CTCAGTCAGCCGCTCAACACCGCTGGACGCC----- 867
150 ArgLysAsnAspMetThrIleLeuPheAlaGlyAsnGluGlyProAsnGlyGlyThr 169
868 -----GGTATAGTAGTCTCGGTGCGCGCGGCAACACGCGGCGCAACCTACACC 918
170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg 189
919 GTCGGCTCACCGCGCGGAGAGTCAACCGTGGTGA----- 963
190 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 209
964 -----GTTGACGAGCAACCAACATCGCGAGCTTCTCCAGCAGGGA 1005
210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSer 229
1006 CCAGCCGCGGAGGAGGCTCAACCGGAGTGTGCGCGCGGCTTCACATCATAGCC 1065
230 AlaArgSerSerLeuAlaProAspSerPheTrpAlaAsnHisAspSerLysTyrAla 249
1066 CCGCGCGCCAGC-----GGAACACGAGTGGCCACCGCGGATAAAGACTACTACACC 1116
250 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly---AsnValAlaGlnLeu 268
1117 AAGCCCTCTGGAACCATGCGCCCGGAGTGTGCGCGGCTTGGCGCGCTCATCCTC 1176
269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeuLys 286
1177 CAGGCCAC-----CGAGTGGACCCCGGACCAAGGTGAAG 1212
287 AlaAlaLeuIleAlaGlyAla-----AlaAspIleGlyLeu 298
1213 ACCGCCCTCTGAGACCGCGGACATAGTGGCCCGGAGGATAGCGGACATCGCTAC 1272
299 GlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeu----- 316
1273 GGTGCG-----GGTAGGGTCAACGCTTACAAAGGCCATCAAGTAC 1311
317 ---AsnValAlaTyrValAsnGluSerSerLeuSerThrSerGlnLysAlaThrTyr 335
1312 GACGACTACGCAAGCTCACTTACCGGCTCCGTCGCGGACAAAGGAGCGGCCACCCAC 1371
336 SerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaPro 355
1372 ACCTTCGACGTCAGCGGCGCCACCTTCGTGACCGCCACCTCTACTGCGGAC----- 1422
356 AlaSerThrThrAlaSerValThrLeuValAsnAspLeuAspLeuValIleThrAlaPro 375
1423 -----ACGGGCTCGAGCGACATCGACCTCTACCTCTACGACCCC 1461
376 AsnGlyThrGlnTyrValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGly 395
1462 AACGGGAACGAG-----GTTGACTACTCTTACACCGCTACTAC----- 1500
396 ArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrIleGlu 415
1501 -----GGCTTCGAGAAGTCCGGCTACTACAACCGCGCGGCAACCTCGAGCGTCAAG 1554
416 ValGlnAlaTyrAsnValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
1555 GTCGTCAGCTACAAG-----GGCGCGGCGAAGTACACGAGTGCAGCTCGTCAGC 1602

RESULT 6
US-09-445-472-11
Sequence 11, Application US/09445472
Patent No. 6358726
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: SHIMOJO, Tomoko

Alignment Scores:
Pred. No.: 9,03e-37 Length: 1977
Score: 452.50 Matches: 138
Percent Similarity: 44.66% Conservative: 67
Best Local Similarity: 30.07% Mismatches: 153
Query Match: 20.14% Indels: 101
DB: 4 Gaps: 18

US-09-985-689A-1 (1-434) x US-09-445-472-11 (1-1977)

QY 8 ValLysAlaAspValAlaGlnSerSerTyrGlyGlnGlyGlnIleValAla 27
Db 433 ATAGGGCGGATACCGTCTGGAACTCCCTCGGCTAGCAGCGAGCGGTGGTGGTGC 492
QY 28 ValAlaAspThrGlyLeuAsnThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
Db 493 ATGTCGTATACGGGTATAGACGCGAAC-----CACCCCGATCTGAAG 534
QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn 66
Db 535 GGCAAGGTCTATAGGCTGTGACGACGCGGTACCGGAGGTGCGACCGCTACGATGACCCAG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLys----- 84
Db 595 GGACACGGAACCCACGCTTCGGGTATCGTTCGCGAACCGCGCAGCGTTAACTCCCGATC 654
QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 101
Db 655 ATAGCGGTGCGCGCGCGGCGGAGCTCGTCGCGGTCAAGGTCTCTCGGTGCGGACGCTCG 714
QY 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 121
Db 715 CGAAGCGTCTCCACCATCATCGCGGTGTGTGACTGGTGTGCTCCAGAACAGGACAAAGTAC 774
QY 122 GlyAlaArgIle-----HisThrAsnSer 129
Db 775 GGGATAAGGGTCTACCACTCTCCCTCGGCTCTCCAGAGCTCCGAGCGAACCGACTCC 834
QY 130 TrpGlyAlaAlaValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrVal 149
Db 835 CTCAGTCAGCGCTCAACACGCGCTGGAGCGC----- 867
QY 150 ArgLysAsnAspMetThrIleLeuPheAlaGlyAsnGluGlyProAsnGlyGlyThr 169
Db 868 -----GGTATAGTAGTCTCGGTGCGCGCGGCAACACGCGGCGGCAACCTTCCAGCAGGGA 918
QY 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGlyAsnLeuArg 189
Db 919 GTCGGCTCACCGCGCGGAGGAGTCAACCGTGGTGA----- 963
QY 190 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 209
Db 964 -----GTTGACGAGCAACCAACATCGCGAGCTTCTCCAGCAGGGA 1005
QY 210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSer 229

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1566 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
FEATURE:
OTHER INFORMATION: /note= N at position 1283 is G or T.
US-08-894-818B-4

Alignment Scores:
Pred. No.: 3.17e-33 Length: 1566
Score: 416.50 Matches: 141
Percent Similarity: 42.55% Conservative: 59
Best Local Similarity: 30.00% Mismatches: 148
Query Match: 18.54% Indels: 122
DB: 3 Gaps: 20

US-09-985-689A-1 (1-434) x US-08-894-818B-4 (1-1566)

QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
Db 24 GTCTGCAGCTCAAGTTATGGCAACTTACGTTTGGAACTTGGGATATGATGTTCTGGAAT 83
QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisG1 44
Db 84 CACAATAGGAATAATTGACACTGGAATTGAC-----GCTTCTCATCC 125
QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 126 AGATCTCCAAGGAAAGTA-----ATTGGTGGGTAGATTGTGCAATGG 170
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 171 TAGGAGTTATCATCAGTACGATGACATGGACATGGAACATCATGCTTCAATAGCAGCTGG 230
QY 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 231 TACTGGAGCAGCAAGTAATGGCAAGTACAAAGGAATGGCTCCAGAGCTAAGCTGGCGGG 290
QY 94 eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
Db 291 AATTAAGGTCTAGGTGCGCATGTTCTTGGGAAGCAATATCTACTATAATAGGAGGTGA 350
QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 351 GTGGCCGTTGATACAAAGATAAGTACGGAATTAAGTCAATCTTCTCTTGGTTC 410
QY 132 a-----AlaValAsnGlyAlaTyrThrTh 140
Db 411 AAGCCAGAGCTCAGATGGTACTGACGCTCTAAGTCAAGCTGCTTAATGAGCGGTGGATGC 470
QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 471 T-----GGATTAGTTGTGTGTTGCCGC 494
QY 160 aGlyAsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaAl 180
Db 495 TGGAAACAGTGGACCTAACCAAGTATACAATCGTTCTCCAGCAGCTGCAAGCAAGTTAT 554
QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 555 TACAGTTGGAGCC-----GTTGACAGTATGA 581
QY 200 nHisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 582 TGTATTAACAAGCTTCTCAAGCAGAGGCCCAACTGCAGACGCGAGCTTAAGCCTGAGT 641
QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPh 240

294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgVa 310
Db 852 AGATGAATAGCGGATATACCTACGGTGCA-----GGTAGGTT 890
QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSerThrSe 330
Db 891 TAATGCATACAGGCTATAAAC-----TAGGTAACCTATGCAAGCTAGTGTTCACCTGG 944
QY 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProle 345
Db 945 ATATGTGGTCAACAAAGGCGAGCCAACTCACCAGTTGTTATTACGGAGCTTCGTTCT 1004
QY 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
Db 1005 AACTGCCACATTTACTGGGCAATGCCAAT----- 1035
QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
Db 1036 -AGCGACCTTGATCTTTACCTCTACGATCCCAATGGAAACCAG---GTTGACTACTCTTA 1091
QY 385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1092 CACCGCTACTAT-----GGATTGGAAGAGGTTGGTTATTATTA 1127
QY 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1128 CAACCCCACTGATGGAACATGGACAATTAAGTTGTTAAGCTACAGC-----GGAAG 1178
QY 425 oGlnThrPheSerLeuAlaIleValAsn 434
Db 1179 TGCNAATCACTCAAGTAGATGTTGGTAAAGT 1206

RESULT 8
US-08-894-818B-4
Sequence 4, Application US/08894818B
Patent No. 6261822
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MITTA, Masanori
APPLICANT: ASADA, Kiyozo
APPLICANT: TSUNASAWA, Susumu
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA=1

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894.818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1962 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-894-818B-34
;
; Alignment Scores:
; Pred. No.: 4,52e-33 Length: 1962
; Score: 416.50 Matches: 141
; Percent Similarity: 42.55% Conservative: 59
; Best Local Similarity: 30.00% Mismatches: 148
; Query Match: 18.54% Indels: 122
; DB: Gaps: 20
;
; US-09-985-689A-1 (1-434) x US-08-894-818B-34 (1-1962)
;
; QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
; DB 420 GTCGACGCTCAAGTTATGCGACACITACGTTGGAACTGGGATATGATGTTCTGGAAT 479
; QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisG1 44
; DB 480 CACAATAGGAATAATTGACACTGGAATTGAC-----GCTTCTCATCC 521
; QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
; DB 522 AGATCTCCAAGAAAAGTA-----ATTGGTGGGTAGATTTCATATGG 566
; QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
; DB 567 TAGGAGTTATCCATACGATGACCATGACATGGAACATCATGTAGCTTCAATACAGCTGG 626
; QY 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
; DB 627 TACTGGAGCAGCAAGTAATGCAAGTAAGGAAATGGCTCCAGGAGCTAAGCTGCGGG 686
; QY 94 eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
; DB 687 AATTAAAGTTCTAGTGCGCGATGTTCTTGGAGCATATCTACTATAATTAAGGAGTTGA 746
; QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
; DB 747 GTGGGCGGTTGTGATAACAAAGATAAGTACGGAATTAAAGGTCATTAACTTCTTCTGTTTC 806
; QY 132 a-----AlaValAsnGlyAlaTyrThrTh 140
; DB 807 AAGCCAGCTCAGATGGTACTGACGCTCTAAGTCAGGCTGTATATGACGCGTGGATGC 866
; QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAla 160
; DB 867 T-----GGATTAGTTGTTGGTGGCCGC 890
; QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla 180

```

RESULT 9

```

; US-08-894-818B-34
; Sequence 34, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozi
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```

```

Db      891  TGGAAACAGTGGACCTAACAAAGTATACAAATCGGTCTCCAGCAGCTGCAACAAAGTTAT 950
Qy      180  eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerThrAlaAspAsnIleAs 200
Db      951  TACAGTTGGAGCC-----GTTGACAAAGTATGA 977
Qy      200  nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db      978  TGTATATAACAAAGCTTCTCAACGACAGAGGCCAACTGCGAGCGCAGGCTTAAGCCCTGAGGT 1037
Qy      220  lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db      1038  TGTGTCTCCAGAACTGGGATAATGTGTCAGAGCAAGT-----GGAAGTACGAT 1088
Qy      240  eTrpAlaAsnHisAspSerLysThrAlaTyMetGlyGlyThrSerMetAlaThrProI 260
Db      1089  GGTCAACCAATTAATGACTATTACACAGCAGCTCTGGGACATCAATGGCAACTCCCTCA 1148
Qy      260  eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db      1149  CGTAGCTGGATTGCGACCCCTCTGTCTCCAA-----GCACACCC 1187
Qy      280  oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
Db      1188  GAGCTGGACTCCACACAAAGTAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 1247
Qy      294  -----AlaAspIleGlyLeuGlyTyProAsnGlyAsnGlnGlyTrpGlyArgVa 310
Db      1248  AGATGAATAGCCGATATAGCTACGGTGCA-----GGTAGGGT 1286
Qy      310  lThrLeuAspLysSerLeuAsnValAlaTyValAsnGluSerSerSerLeuSerThrSe 330
Db      1287  TAATGCATACAGGCTATAAC-----TACGATAACTATGCAAGCTAGTGTCTCACTGG 1340
Qy      330  rGlnLysAla-----ThrTySerPheThrAlaThrAlaGlyLysProLe 345
Db      1341  ATATGTTGCCAAACAAAGCCAGCCAACTCACCAGTTGCTATTATAGCGAGCTTCGTCGT 1400
Qy      345  uLysIleSerLeuValTrpSerAspAlaProAlaSerThrAlaSerValThrLeuVa 365
Db      1401  AACTGCCACATTAATACCTGGGCAATGCCAAT----- 1431
Qy      365  lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyValGlyAsnAspPh 385
Db      1432  -AGCGACTTGATCTTTACCTCTACGATCCCAATGGAACCCAG---GTTGACTACTCTTA 1487
Qy      385  eThrSerProTyAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db      1488  CACCGCTACTAT-----GGATTCGAAAAGGTGGTTATTA 1523
Qy      405  nAlaProGlnSerGlyThrTyThrIleGluValGlnAlaTyAsnValProValGlyPr 425
Db      1524  CAACCCAACTGATGAACATGGCAATTAAGTTGTAAGCTACAGC-----GGAAG 1574
Qy      425  oGlnThrPheSerLeuAlaIleValAsn 434
Db      1575  TCGAAACTATCAAGTAGATGTGGTAAGT 1602

RESULT 10
US-09-445-472-15
; Sequence 15, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06

```

```

; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 15
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-15

Alignment Scores:
Pred. No.: 4,52e-33 Length: 1962
Score: 416.50 Matches: 141
Percent Similarity: 42.55% Conservative: 59
Best Local Similarity: 30.00% Mismatches: 148
Query Match: 18.54% Indels: 122
DB: 4 Gaps: 20

US-09-985-689A-1 (1-434) x US-09-445-472-15 (1-1962)
Qy      12  ValAlaGlnSerSerTyGly-LeuTy-----GlyGlnGlyGl 24
Db      420  GTCTGCAGCTCAAGTTATGGCAACTTACGTTTGGAACTTGGATATGATGTTCTGGAAT 479
Qy      24  nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGl 44
Db      480  CACAATAGGAATAATTGACACTGGAATTGAC-----GCTTCTCATCC 521
Qy      44  uAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db      522  AGATCTCCAAGGAAAGTA-----ATTGGTGGGTAGATTGTCATATGG 566
Qy      64  -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGl 78
Db      567  TAGGAGTTATCCATACGATGACCATGGACATCGAATCATGTAGCTTCAATAGCAGCTGG 626
Qy      78  yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db      627  TACTGGACGACGCAAGTAATGGCAAGTACAAAGGAATGGCTCCAGGAGCTAAGCTGGCGGG 686
Qy      94  eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyLeuProSerAsnLeuGl 112
Db      687  AATTAAGTTCTAGTGCGCATGTTCTGGAACCATATCTACTATATTAAGGAGGTGA 746
Qy      112  nThrLeuPheSerGlnAlaTySerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db      747  GTGGCCCGTGTATAACAAAGATAAGTACGGAATTAAGGTCAATTAATCTTCTCTGTGTC 806
Qy      132  a-----AlaValAsnGlyAlaTyThrTh 140
Db      807  AAGCCAGAGCTCAGATGGTACTGACGCTCTAAGTCAGGCTGTTAATGCAGCTGGGATCC 866
Qy      140  rAspSerArgAsnValAspTyValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db      867  T-----GGATTAGTTGTTGTTGGTGGCGC 890
Qy      160  aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIl 180
Db      891  TGGAAACAGTGGACCTTAACAAGTATACAATCGGTTCTCCAGCAGCTGCAACAAAGTTAT 950
Qy      180  eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAs 200
Db      951  TACAGTTGGAGCC-----GTTGACAAAGTATGA 977
Qy      200  nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db      978  TGTATATAACAAAGCTTCTCAACGACAGAGGCCAACTGCGAGCGCAGGCTTAAGCCCTGAGGT 1037
Qy      220  lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db      1038  TGTGTCTCCAGAACTGGGATAATGTGTCAGAGCAAGT-----GGAAGTACGAT 1088

```

```

QY 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProI 260
Db 1089 GGGTCAACCAATTAATGACTATTACACAGCAGCTCTGGGACATCAATGGCAACCTCTCA 1148
QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyLeuThrPr 280
Db 1149 CGTAGCTGGTATTCAGCCCTTTCCTCAA-----GCACACCC 1187
QY 280 oLys-----ProSerLeuLeuLysAlaAlaLeuLeuAlaGlyAla----- 293
Db 1188 GAGCTGAGCTCCAGCACAAGTAAACACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 1247
QY 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrPArgVa 310
Db 1248 AGATGAATAAGCCGATATAGCTACGGTGCA-----GGTAGGGT 1286
QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSerThrSe 330
Db 1287 TAATGCATACAGGCTATAAAC-----TAGGATAACTATGCAAGAGCTAGTGTCTACTGG 1340
QY 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
Db 1341 ATATGTTGCCAACAAAGCCAGCCAACTCACCAGTTCGTTATTAGCGGAGCTTCGTTGCT 1400
QY 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrAlaSerValThrLeuVa 365
Db 1401 AACTGCCACATTTACTTGGGACCAATGCCAAT----- 1431
QY 365 lAsnAspLeuAspValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
Db 1432 -AGCGACCTTGATCTTTACCTCTACGATCCCAATGGAACACAG---GTTGACTACTCTTA 1487
QY 385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1488 CACCGCCTACTAT-----GGATTGCAAGAGTTGTTATT 1523
QY 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1524 CAACCCAACTGATGGAAACATGGACAATTAAGGTTGAAGCTACAGC-----GGAAG 1574
QY 425 oGlnThrPheSerLeuAlaIleValAsn 434
Db 1575 TGCAAACTATCAAGTAGATGTGGTAAGT 1602

```

RESULT 11

```

US-08-894-818B-6
; Sequence 6, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B

```

```

; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1977 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-894-818B-6

```

```

Alignment Scores:
Pred. No.: 9,95e-32 Length: 1977
Score: 403.50 Matches: 134
Percent Similarity: 43.07% Conservative: 68
Best Local Similarity: 28.57% Mismatches: 147
Query Match: 17.96% Indels: 120
DB: 3 Gaps: 20

```

US-09-985-689A-1 (1-434) x US-08-894-818B-6 (1-1977)

```

QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
Db 420 GTCTCGACGCTCAAGTTATGGCAACTTACGTTTGGAACTTGGGATATGATGGTCTGGAAT 479
QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisG1 44
Db 480 CACATAGAATAATTGACACTGGAATTGAC-----GCTTCTCATCC 521
QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCAAGAAAGTA-----ATTGGGTGGGTAGATTTCGTCATCG 566
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 567 TAGGAGTTATCCATACGATGACCATGGACATGGAATCATAGCTTCAATAGCAGCTGG 626
QY 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAATGGCAAGTACAAAGGAATGGCTCCAGGAGCTAAGCTGGCGGG 686
QY 94 eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
Db 687 AATTAAGGTCTTAGTGCCGATGGTTCTGGAAGCATATCTACTATAATTAAGGAGTTGA 746
QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIle----- 125
Db 747 GTGGGCGGTTGATAACAAAGATAAGTACGGAATTAAGGTCATTAATCTTCTTGTGTTTC 806
QY 126 -----HisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrTh 140
Db 807 AAGCCAGAGCTCCGACGACCGACTCCCTCAGTCAGCGCGTCAACACCGCTGGGAGCG 866
QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaI 160
Db 867 C-----GGTATAGTAGTCTCGCTGCCGCGC 890
QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI 180
Db 891 CGGCAACAGCGGCGGCAACACCTACACCGTGGCTCACCCGCGCGCGGCAAGGTCAT 950

```


Db 1286 GCGCCGCGCGAGCGCGCGTCTGTCACCATGACGCTGGGC-----GGCATG 1333
QY 138 TyrThrThrAspSerArgAsnValAspAspTyrValArgLys-----AsnAspMetThr 155
Db 1334 GACACCGGAGACCGCCCTGGAGCGCGCGGTTCGACAGCTGCTCCGCGAGAGAGCGC 1393
QY 156 IleLeuPhe-----AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaPro 173
Db 1394 GTCCTGTTCGTCATCGCGCGCGCGCAACGAGGCGCCGAG-----TCGATCGGTTCGCC 1447
QY 174 GlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGly 193
Db 1448 GGCAGCGCGGAGCGCGCTCACCGCTCGCGCGC----- 1480
QY 194 SerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLys 212
Db 1481 -----GTCGACGACGAGGACAGCTCGCGACTTCCTCCACGCGCGCCCGCTCGGC 1534
QY 213 AspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSer 232
Db 1535 GACGGCGCATCAAGCGCGAGCTCACCGCTCCCGCGTGGACATCACGCGCGCTCGCGC 1594
QY 233 SerLeuAlaProAspSerSerPheThrAlaAsnHisAspSerLysTyrAlaTyrMetGly 252
Db 1595 GAGGGAACGACATCGCGCCAGGAGTCTGTCAGGACCGCGCGGTACATGACCATCTCC 1654
QY 253 GlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPhe 272
Db 1655 GGCACGTCGATCGCGACCGCGCGCTCGCGCGCGCGCGCTCTCTGAGCAGCAG--- 1711
QY 273 ValLysAsnArgGlyIleThrProLysProLysProSerLeuLeuLysAlaAlaLeu--- 289
Db 1712 -----CACCCGACTGGACCTCCGCGCAACTGAAGGGCGCG 1747
QY 290 IleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArg 309
Db 1748 CTCACCGGCTCCACCAAG---GGCGCAAGTACACCCCGTTCGAGCAGGGTTCGGGCGG 1804
QY 310 ValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSer--- 328
Db 1805 ATCCAGCGCGAGCAAGCGCTCCAGCAGCGTGTATCGCGACCCCGTTCGAGCAGGGTTC 1864
QY 329 -----ThrSerGlnLysAlaThrTyrSer 336
Db 1865 GCGTCCACGAGTGGCGCGCACACCGACGACGCGCGCCACACGAGTGCCTACCTACCGC 1924
QY 337 PheThrAlaThrAlaGlyLysProLysProLysIleSerLeuValTrpSerAsp----- 353
Db 1925 AACCTCGGCACCCAGCAGCTCAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 1984
QY 354 -----AlaProAlaSer 357
Db 1985 AAGCGCGCGCGCGCGCTTCTTACGCTGGCGCGCCACACGAGTGCCTCCGCGCGCG 2044
QY 358 ThrThrAlaSerValThrLeuValAsnAspLeuAspLeu----- 370
Db 2045 GGCAGCGCTCGTGCATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2104
QY 371 -----ValIleThrAlaProAsnGlyThrGlnTyrVal 381
Db 2105 TACTCGCGCTAGTGTGCGCCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2146

RESULT 13

US-09-514-340-3
; Sequence 3, Application US/09514340
; Patent No. 6361987
; GENERAL INFORMATION:
; APPLICANT: AKIRA ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROXYRINDINE DERIVATIVES
; ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/514,340
FILING DATE: 28-Feb-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/000,016
FILING DATE: January 30, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Lee Cheng
REGISTRATION NUMBER: 40,949
REFERENCE/DOCKET NUMBER: <Unknown>
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX: <Unknown>

SEQUENCE CHARACTERISTICS:
LENGTH: 2539 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces viridosporus
STRAIN: A-914
FEATURE:

NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
SEQUENCE DESCRIPTION: SEQ ID NO: 3;
US-09-514-340-3

Alignment Scores:
Pred No.: 3,728-26 Length: 2539
Score: 351.00 Matches: 131
Percent Similarity: 42.63% Conservative: 54
Best Local Similarity: 30.18% Mismatches: 159
Query Match: 15.62% Indels: 90
DB: 4 Gaps: 16

US-09-985-689A-1 (1-434) x US-09-514-340-3 (1-2539)

QY 2 AspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
Db 953 GACAGCTCGTGGCGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1009
QY 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 41
Db 1010 AAGGCGGTGAAGATCGCGCTCTGGACACCGGTGTGCACACGAGC----- 1054
QY 42 MethisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCGGACCTGAAGGCGCGGTGACCGGTCCCAAGACTTCACCGCGCGCGCGCG 1111
QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySer 81
Db 1112 GCGCGCGACAAAGTGGCGCACCGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1171
QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
Db 1172 CAGTCCAAGGCAAGTACAGGCGGTCTGACCGCGCGCGCGCGCGCGCGCGCGCGCG 1231

QY 42 MethHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCGGACCTGAAGGGCCGGTGACCGCGTCCAAAGAACTTCAACCGCGCGCCGCC 1111
QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySer 81
Db 1112 GCGGCGCAGAAAGTGGCGCACCGGACCGACCTCGCTCATCGCGGCGGCGACGGCGCC 1171
QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
Db 1172 CAGTCCAAAGGCAAGTACAGGGCGTGCACCGCGCGCGATCTCAACGCGGAAGGTC 1231
QY 98 MetAspSerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
Db 1232 CTCGAGACATCCGGT-----TTCGGCGACGACTCCGCGATCTCCGCGGATGGAGTGG 1285
QY 118 AlaTyrSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAlaValAsnGlyAl 137
Db 1286 GCGGCGCGCAGGCGCGGACGTCGTCAATGAGCTGGCGGATGGACACACCGGAG 1345
QY 137 aTyrThrThrAspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLe 157
Db 1346 ACCGACCGCTGAGGCGCGG-GTCCGACAGCTCTCCCGCAGAGGCGCTCTGTTCCG 1404
QY 157 uPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLy 177
Db 1405 CATCGCGCGCGGCAACGAGGCGCGGAG-----TCGATCGTTCGCGCGCAGCGCGA 1458
QY 177 sAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAs 197
Db 1459 CGCGCGCTCACGTCGGCGC-----GTCTGA 1485
QY 197 pAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLysAspGlyArgI 216
Db 1486 GCACAAAGGACAAGCTCGCGGACTTCTCTCCACCGCGCGCGCTCGCGCAGCGCGCAT 1545
QY 216 elysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPr 236
Db 1546 CAAGCGGACGTCACCGCTCCCGCGGTGGACATCACGCGCGCTCGCGGAGGCGCAACGA 1605
QY 236 oAspSerSerPheThrAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMe 256
Db 1606 CATCGCGCAGGAGTGGTGAGGACCGCGCGCTACATCACCATCTCCGCGCAGCTCGAT 1665
QY 256 tAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnAr 276
Db 1666 GCGGACCGCGCAGCTCGCGGCGCGCGCTCTCTGAAGCAGCAG----- 1711
QY 276 gGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu-----IleAlaGlyAl 293
Db 1712 -----CACCCCGACTGGACCTCCCGCACTGAAGGCGCGCTCACCGGCTC 1758
QY 293 aAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAs 313
Db 1759 CACCAAG---GGCGGCAAGTACACCCCGTTCGAGCAGGTTCCGGCGGATCCAGGCCGA 1815
QY 313 pLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSer----- 328
Db 1816 CAAGGCGTCCAGCAGACCGTATCCCGACCGCGCTCTCGGTGAGCTTCGGGCTCCAGCA 1875
QY 329 -----ThrSerGlnLysAlaThrTyrSerPheThrAlaTh 340
Db 1876 GTGGCGGCACACCGACGAGCGCGGTCAACCAAGCAGCTGACCTACCGCAACCTCGGCAC 1935
QY 340 rAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp----- 353
Db 1936 CCAGGACGTCACGCTGAAGCTGACGTGACCGCGCACCGCGCCAAAGGCGAAGCGCGCC 1995
QY 354 -----AlaProAlaSerThrThrAlaSe 361
Db 1996 GCGGGGCTTCTTACGCTGGCGGCCACCGGTGACCGTCCCGCGGCGCGGCGCGCTC 2055
QY 361 rValThrLeuValAsnAspLeuAspLeu-----Va 371

Db 2056 CGTCGACATACCGCGCGCACACCGCGCTCGCGCGCAGCGTGGACGGCGTACTCGCGCTA 2115
QY 371 lIleThrAlaProAsnGlyThrGlnTyrVal 381
Db 2116 CGTGTGTCGCCACGGCGCGCGGCGGCGAGCGGTC 2146
RESULT 15
US-09-514-340-1
; Sequence 1, Application US/09514340
; Patent No. 6361987
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPRYDINE
; ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/514,340
; FILING DATE: 28-Feb-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/000,016
; FILING DATE: January 30, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee Cheng
; REGISTRATION NUMBER: 40,949
; REFERENCE/DOCKET NUMBER: <Unknown>
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2809 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces antibioticus
; STRAIN: <Unknown>
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 338...2539
; IDENTIFICATION METHOD: E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2540...2809
; IDENTIFICATION METHOD: P
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-514-340-1
Alignment Scores:
Pred. No.: 1,43e-25 Length: 2809
Score: 346.00 Matches: 129
Percent Similarity: 42.00% Conservative: 52
Best Local Similarity: 29.93% Mismatches: 167
Query Match: 15.40% Indels: 84
DB: 4 Gaps: 13

US-09-985-689A-1 (1-434) x US-09-514-340-1 (1-2809)	
QY	2 AspValAlaArgGlyLeuValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
Db	953 GACAGTCGTCGGCGAGATCGCGCCGCCCAAGCGTGGTCCGCC---GGCTACGACGGC 1009
QY	22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSer 41
Db	1010 AAGGCGTGAAGATCGCTCTCGACACCGGTGTCACACGAGC----- 1054
QY	42 MethHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
Db	1055 ---CATCCGGAGCTGAAGCGCGGTGACCGCGTCCAGAACTCACCGCGCGCCCGGC 1111
QY	62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySer 81
Db	1112 GCCGCGACAAAGTGGCGCCACGCCACCGCTCGCATCGCGGGCGGCAGCGGCC 1171
QY	82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
Db	1172 CAGTCCCAAGCGCAAGTACAGGCGGTGCGACCGCGCGCGGATCCTCAACGGCAAGTGC 1231
QY	98 MetAspSerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
Db	1232 CTCGACGACTCCGCT-----TTCGGCGACGACTCCGGCATCCTCGCGCGCATGGAGTGG 1285
QY	118 AlaTyrSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAlaValAsnGlyAl 137
Db	1286 GCGCGCGCGAGCGCGCGCGAGTCTGTCACATGAGCCTGGCGCGCATGGACACCGGAG 1345
QY	137 aTyrThrThrAspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleIle 157
Db	1346 ACCGACCCGCTGAGCGGGCGG-GTCGACAGCTCTCCGCCGAGAGGGCGTCTGTTCGC 1404
QY	157 uPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLy 177
Db	1405 CATCGCGCGCGCAACGAGGCGCGGAG-----TCGATCGGTTCGCCCGCGAGCGCGGA 1458
QY	177 sAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAs 197
Db	1459 CGCGCGCTCACCCTCGCGGCC-----GTGCA 1485
QY	197 pAsnIleAsnHisValAlaGlnPheSerArgGlyPro---ThrLysAspGlyArgIle 216
Db	1486 CGACAAGGACAAGTCCCGACTTCTCTCCACCGCGCGCGCTCGCGCGAGCGGGCCAT 1545
QY	216 eLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPr 236
Db	1546 CAAGCGGAGTCCACCGCTCCCGCGTGGAGCATCACGGCGGCTCGCGGGAGGGGNACGA 1605
QY	236 oAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMe 256
Db	1606 CATCGGCGAGAGTCGGTGAGGACCGCGCGCTACATGACCATCTCCGGCGACGTCGAT 1665
QY	256 tAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnAr 276
Db	1666 GCGCACCCCGCACGTCGCGGGCGCGCGCTCTCTGAAGCAGCAG----- 1711
QY	276 gGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu-----IleAlaGlyAl 293
Db	1712 -----CACCCCGACTGACCTCCCGCGCACTGAAGGCGCGCTCACCGGCTC 1758
QY	293 aAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAs 313
Db	1759 CACCAAG---GGCGCAAGTACACCCCGTTCGAGCAGGTTTCGGGCGCGATCCAGGCCGA 1815
QY	313 pLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSer----- 328
Db	1816 CAAGGCGCTCCAGCAGACCGTATCGCGCGCGGCTCTCGGTGAGCTTCGGCGTCCAGCA 1875
QY	329 -----ThrSerGlnLysAlaThrTyrSerPheThrAlaTh 340
Db	1876 GTGGCCGCACACCGAGCAGCGCGTCTACCAAGCAGCTACCTACCGCAACCTCGGCAC 1935

QY	340 rAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp----- 353
Db	1936 CCAGGACGTCACGCTGAAGCTGACGTCCAGCGCCACCGCCCAAGGCGGCGCCC 1995
QY	354 -----AlaProAlaSerThrThrAlase 361
Db	1996 GCGGGCTTCTTCACGCTGGGGCGCCACCACGGTGACCGCTCCCGCGGCGGCGGCGCTC 2055
QY	361 rValThrLeuValAsnAspLeu-----Va 371
Db	2056 CCGTCGACATGACCGCGGACACCGCGCTCGGGCGGCGGCGGCGGCGGCGGCTC 2146
QY	371 lIleThrAlaProAsnGlyThrGlnTyrVal 381
Db	2116 CGTGGTCCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTC 2115

Search completed: July 28, 2003, 08:42:47
Job time : 131.115 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein : nucleic search, using frame_plus_p2n model

Run on: July 28, 2003, 03:48:55 ; Search time 332.884 Seconds
(without alignments)
2689.657 Million cell updates/sec

Title: US-09-985-689A-1

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVOAYNPVGPQTFLAIVN 434

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_p2n.model -DEV=xlp
-O=/cgn2_1/USPRO_pool/US09985689/tunat_25072003_143033_892/app_query.fasta_1.1166
-DB=Published_Applications_NA -OFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -STARF=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DCCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09985689@cgn_1_1_409.@runat_25072003_143033_892
-NCPU=6 -ICPU=3 -NO_MMWP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*

3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*

7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*

10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*

11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*

12: /cgn2_6/ptodata/2/pubpna/US09D_PUBCOMB.seq:*

13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*

14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*

15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*

16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*

17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	452.5	20.1	1977	13	US-10-090-624-11 Sequence 11, Appl

2	416.5	18.5	1236	13	US-10-090-624-11 Sequence 13, Appli
3	416.5	18.5	1962	13	US-10-090-624-15 Sequence 15, Appl
4	363.5	16.2	3624	14	US-10-156-761-15 Sequence 5701, Ap
5	363.5	16.2	9025608	14	US-10-156-761-15 Sequence 1, Appli
6	346.5	15.4	3417	14	US-10-156-761-3306 Sequence 3306, Ap
7	346.5	15.4	9025608	14	US-10-156-761-3306 Sequence 1, Appli
8	332.5	14.8	1329	10	US-09-974-300-1934 Sequence 1934, Ap
9	307	13.7	4765	13	US-10-090-624-5 Sequence 5, Appli
10	283	12.6	3743	11	US-09-927-827-29 Sequence 29, Appl
11	282	12.6	3788	11	US-09-927-827-33 Sequence 33, Appl
12	276	12.3	1306	10	US-09-966-921A-5 Sequence 5, Appli
13	276	12.3	1330	10	US-09-966-921A-5 Sequence 5, Appli
14	270	12.0	3303	14	US-10-156-761-5384 Sequence 5384, Ap
15	256	11.4	1485	10	US-09-974-300-1938 Sequence 1938, Ap
16	253	11.3	1971	10	US-09-974-300-1935 Sequence 1935, Ap
17	251	11.2	840	14	US-10-209-812-1 Sequence 1, Appli
18	242.5	10.8	3884	11	US-09-927-827-34 Sequence 34, Appl
19	240	10.7	1140	8	US-08-322-678-11 Sequence 11, Appl
20	240	10.7	1140	8	US-08-322-678-12 Sequence 12, Appl
21	237.5	10.6	1497	8	US-08-322-678-6 Sequence 6, Appli
22	237.5	10.6	1497	9	US-09-060-854B-1 Sequence 1, Appli
23	237.5	10.6	1497	14	US-10-033-325-1 Sequence 1, Appli
24	237.5	10.6	1497	14	US-10-228-572-1 Sequence 1, Appli
25	235.5	10.5	1494	14	US-10-104-693-1 Sequence 30, Appl
26	232	10.3	3452	11	US-09-927-827-30 Sequence 5689, Ap
27	231.5	10.3	1332	14	US-10-156-761-5689 Sequence 13, Appl
28	227.5	10.1	1140	10	US-09-920-118-13 Sequence 5, Appli
29	226.5	10.1	1457	14	US-10-202-339-1 Sequence 3, Appli
30	225.5	10.0	4198	9	US-09-891-711-5 Sequence 14, Appl
31	225.5	10.0	4338	9	US-09-837-235-14 Sequence 1, Appli
32	223.5	9.9	1074	9	US-10-328-459-1 Sequence 17, Appl
33	223.5	9.9	2760	14	US-09-824-893A-17 Sequence 14, Appl
34	222	9.8	522	11	US-09-824-893A-14 Sequence 15, Appl
35	220	9.7	1146	10	US-10-156-761-2788 Sequence 2788, Ap
36	218.5	9.7	1206	11	US-09-824-893A-12 Sequence 112, App
37	218.5	9.7	1206	11	US-09-824-893A-15 Sequence 15, Appl
38	218	9.7	522	11	US-09-824-893A-9 Sequence 9, Appli
39	215	9.6	522	11	US-09-824-893A-12 Sequence 12, Appl
40	214	9.5	522	11	US-09-824-893A-7 Sequence 7, Appli
41	213	9.5	522	11	US-09-824-893A-9 Sequence 9, Appli
42	213	9.5	522	11	US-09-824-893A-12 Sequence 12, Appl
43	213	9.5	522	11	US-09-824-893A-13 Sequence 13, Appl
44	213	9.5	522	11	US-09-824-893A-13 Sequence 13, Appl
45	213	9.5	522	11	US-09-824-893A-13 Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-10-090-624-11
; Sequence 11, Application US/10090624
; Publication No. US2002013235A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090, 624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 11
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:

PROV DP
NO-THIS DNA

OTHER INFORMATION: Synthetic
US-10-090-624-11

Alignment Scores:

Pred. No.: 2,42e-40 Length: 1977
Score: 452.50 Matches: 138
Percent Similarity: 44.66% Conservative: 67
Best Local Similarity: 30.07% Mismatches: 153
Query Match: 20.14% Indels: 101
DB: 13 Gaps: 18

US-09-985-689a-1 (1-434) x US-10-090-624-11 (1-1977)

QY	8	ValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnIleValAla	27
DB	433	ATAGGGCCGATACCGTCTGGAACCTCCCTCGCTACGACGACGCGTGTGGTGGTGC	492
QY	28	ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg	47
DB	493	ATCGTGCATACGGGTATAGACCGAAC-----CACCCCGATCTGAAG	534
QY	48	GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn	66
DB	535	GGCAAGGTCATAGGCTGTGACGACGCGTCAACGCGAGTCCGACCCCTACGATGACCCAG	594
QY	67	GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLys-----	84
DB	595	GGACACGGAACCCAGTTCGGGTATCGTTCGGGACCGGACCGGACGCTTAACCTCCAGTAC	654
QY	85	---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly	101
DB	655	ATAGCGCTCGCCCGCGCGAAGCTCGTTCGGGTCAAGGTCTCGGTGCGCGGTTCG	714
QY	102	GlyGlyLeuGlyGlyProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla	121
DB	715	GGAAAGCTCTCCACCATCATCGCGGTGTGTGAGTGGTCTCCAGAACAGGACAGTAC	774
QY	122	GlyAlaArgIle-----HisThrAsnSer	129
DB	775	GGATAAGGTCATCAACCTCTCCCTCGGCTCTCCAGAGCTCCGACGGAACGACTCC	834
QY	130	TrpGlyAlaAlaValAsnGlyAlaTyrThrThrAspSerArgAsnValAspTyrVal	149
DB	835	CTCAGTCAGCGCGTCAACACGCTCGGAGCC-----	867
QY	150	ArgLysAsnAspMetThrIleLeuPheAlaGlyAsnGluGlyProAsnGlyGlyThr	169
DB	868	-----GGTATAGTCTCGTTCGCGCGCGGCAACAGCGGCGGCAACCTACACC	918
QY	170	IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg	189
DB	919	GTCGGTCAACCCCGCGCGGAGGTCATACCGTCTCGGTGCA-----	963
QY	190	ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly	209
DB	964	-----GTTGACAGCAACGACATCATCTCCAGCTTCCTCCAGCGGGA	1005
QY	210	ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSer	229
DB	1006	CCGACCGCGGACGGAAGCTCAAGCCGGAAGTCTCGCGCCCGCGGCTTGACATCATGCC	1065
QY	230	AlaArgSerSerLeuAlaProAspSerSerPheThrAlaAsnHisAspSerLysTyrAla	249
DB	1066	CCGCGCGCCAGC-----GGAACGACGATGGCCACCCCGGATAAAGACTACTACACC	1116
QY	250	TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly---AsnValAlaGlnLeu	268
DB	1117	AAGGCTCTGGAACCATGCGCCACCCCGGACGTTTCGGCGGTTCGGCGCTCATCTCTC	1176
QY	269	ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeuLys	286
DB	1177	CAGGCCAC-----CCGAGCTGGACCCCGGACAAAGGTGAAG	1212

QY	287	AlaAlaLeuIleAlaGlyAla-----AlaAspIleGlyLeu	298
DB	1213	ACCGCCTCATCGAGACCGCCACATAGTCGCCCCCAAGGAGATCGGACATCGCCTAC	1272
QY	299	GlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeu-----	316
DB	1273	GGTGG-----GGTAGGGTGAACGCTCTACAAGGCAATCAAGTAC	1311
QY	317	---AsnValAlaTyrValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyr	335
DB	1312	GACGACTACGCCAAGCTCACCTTCACGGCTCGTCCGCCACAGGGAAGGCCACCCAC	1371
QY	336	SerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValThrPheSerAspAlaPro	355
DB	1372	ACCTTCGAGCTCAGCGGCGCCACCTTCGTGACCGCCACCTCTACTGGGAC-----	1422
QY	356	AlaSerThrThrAlaSerValThrLeuValAsnAspLeuAspLeuValIleThrAlaPro	375
DB	1423	-----ACGGGCTCAGGCGACATCGACCTCTACCTCTAGACCC	1461
QY	376	AsnGlyThrGlnTyrValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGly	395
DB	1462	ACGGGAACGAG--GTTGACTACTCTCTACACCGCTACTAC-----	1500
QY	396	ArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrThrIleGlu	415
DB	1501	-----GGCTTCGAGAAGTCTCGGCTACTACACCGGACCGCGGAACTGGACGTCAG	1554
QY	416	ValGlnAlaTyrAsnValProValGlyProGlnThrPheSerLeuAlaIleValAsn	434
DB	1555	GTCGTCAGTACAG-----GGCGCGGGAACCTACACAGTCTCGACGTCGTCAGC	1602

RESULT 2

US-10-090-624-2
Sequence 2, Application US/10090624
Publication No. US20020132335A1
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: SHIMOJO, Tomoko
APPLICANT: ASADA, Kiyoko
APPLICANT: KATO, Kunoshin
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
FILE REFERENCE: TAKAKURA-6
CURRENT APPLICATION NUMBER: US/10/090,624
PRIOR FILING DATE: 2002-03-06
PRIOR APPLICATION NUMBER: 09/445,472
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 151969/1997
PRIOR FILING DATE: 1997-06-10
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 1236
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic
US-10-090-624-2

Alignment Scores:
Pred. No.: 1.25e-36 Length: 1236
Score: 416.50 Matches: 141
Percent Similarity: 42.55% Conservative: 59
Best Local Similarity: 30.00% Mismatches: 148
Query Match: 18.54% Indels: 122
DB: 13 Gaps: 20

US-09-985-689a-1 (1-434) x US-10-090-624-2 (1-1236)

QY	12	ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyGlu	24
DB	24	GTCGACGCTCAAGTTATGGAACCTACGTTTGGAACTTGGGATATGATGTTCTGGAAT	83

Qy	78	yAsnGlySerThrasn	-----LysGlyMetAlaProGlnAlaAsnLeuValPh	94
Db	627	TACTGGAGCAGCAAGTAATGCAAGTACAAAGGAATGGCTCCAGGAGCTAAAGTCGGCGG	686	
Qy	94	eGlnSerIleMet	-----AspSerGlyGlyGlyLeuGlyLeuProSerAsnLeuG1	112
Db	687	AATTAAGGTTCTAGTGGCCGATGGTTCTTGAAGCATATCTACTATATTAATTAAGGAGTGA	746	
Qy	112	nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl	132	
Db	747	GTGGCCGCTTCATAACAAAGATAAGTACGGAATTAAGGTCATTAATCTTCTCTGTGCTTC	806	
Qy	132	a	-----AlaValAsnGlyAlaTyrThrTh	140
Db	807	AAGCGAGAGCTCAGATGGTACTGACGCTCTAAGTCAGGCTGTTAATGCAGCGTGGGATGC	866	
Qy	140	rAspSerArgAsnValAspAspTyrValArgIysAsnAspMetThrIleLeuPheAlaAl	160	
Db	867	T	-----GGATTAGTTCTTGTGGTGGCGC	890
Qy	160	aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaIysAsnAlaI	180	
Db	891	TGGAACAGTGGACCTTAACAAGTATACAAATCGGTTCTCCAGCAGCTCCAAAGTAT	950	
Qy	180	eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs	200	
Db	951	TACAGTTGGAGCC	-----GTTGACAAGTATGA	977
Qy	200	nHisValAlaGlnPheSerSerArgGlyProThrIysAspGlyArgIleLysProAspVa	220	
Db	978	TGTTATAACAAGCTTCTCAAGCAGAGGGCCAACTGCAGAGCGGAGGCTTAAGCTGAGGT	1037	
Qy	220	lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh	240	
Db	1038	TGTTGCTCCAGAAACTGGATAATTGCTGCCAGAGCAAGT	-----GGAACCTAGCAT	1088
Qy	240	eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIl	260	
Db	1089	GGGTCAACCAATTAATGACTATTACACAGCAGCTCTCTGGGCATCAATGGCAACTCCCTCA	1148	
Qy	260	eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIysAsnArgGlyIleThrPr	280	
Db	1149	CGTAGCTGGTATTCGACGCCCTCTGTGTCCAA	-----GCACACC	1187
Qy	280	oLys	-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla	293
Db	1188	GAGCTGGACTCCAGACAAGATAAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCC	1247	
Qy	294	-----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgVa	310	
Db	1248	AGATGAATAGCGCATATAGCTTACGGTGA	-----GGTAGGCT	1286
Qy	310	lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSerThrSe	330	
Db	1287	TAATGCATACAGGCTATAAAC	-----TAGCATACTATCCAAAGCTAGTCTTCACTCG	1340
Qy	330	rglnIysAla	-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe	345
Db	1341	ATATGTTGCCAACAAAGGACGCAAACTCACCAGTTCGTTATTAGCGGAGCTTCGTCGT	1400	
Qy	345	uLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa	365	
Db	1401	AACTGGCCATTTACTGGCAATAGCCAAT	-----	1431
Qy	365	lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh	385	
Db	1432	-AGCGACTTGATCTTTACCTCTACGATCCCAATGGAAACCCAG	-----GTTGACTACTCTTA	1487
Qy	385	eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs	405	
Db	1488	CACCGGCTACTAT	-----GGATTCGAAAGCGTGGTATTATTA	1523

Db 6919455 GTCGACATCCGCGCGCCGCTCCAGCTCGCCCGCGCAGCGCTAC----- 6919502
 QY 245 AspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsn 264
 Db 6919503 -----TACACCTCCATGAGCGGTACGTGATGCGAGCGCCCATGTGCGGGGGTC 6919553
 QY 265 ValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrProLysProSerLeu 284
 Db 6919554 GCGCGCTCCTCCGCGCAGCAGCACCAGCTGACCGCGCGCGCGCTCAAGACCGCTG 6919613
 QY 285 LeuLysAlaAla-----LeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsn 302
 Db 6919614 ATGTCCACGTCCGAGCAACTCGACGCTCCGTATATCAGTTGGGGGGTTCGGTCACT 6919673
 QY 303 GlyAsnGlnGlyTrpGly---ArgValThrLeuAspLysSerLeuAsnValAlaTyr--- 320
 Db 6919674 GTCCGCGAGCGCGCTCGCGCGCGCTCACCGCGCAGCGCGCTCGGTTCAC 6919733
 QY 321 -----ValAsnGluSerSerSerLeuSerThrSerGln 331
 Db 6919734 CGTGGCCCCATCAGCGCGATCAGCCGTCACGAGAGCGGTCACTTCTCCAACTCCTCC 6919793
 QY 332 LysAlaThrTyrSerPheThr---AlaThrAlaGlyLysProLeuLysIleSerLeuVal 350
 Db 6919794 GACACGAGCGTTCGAGTTGAGCCTCGCGGTGCGGGCGCGCGCGGTGTCGCCACCTC 6919853
 QY 351 TrpSerAspAla-----ProAlaSerThrThrAlaSerValThrLeuValAsnAsp 367
 Db 6919854 GCGACACCGCACTCACCGTGGCGCGCCACCGCGCAGCGCGCTGACCGCGCGAC 6919913
 RESULT 6
 US-10-156-761-3306
 ; Sequence 3306, Application US/10156761
 ; Publication No. US200301190181
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIRAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; PRIORITY FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 3306
 ; LENGTH: 3417
 ; TYPE: DNA
 ; ORGANISM: Streptomyces avermitilis
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(3417)
 US-10-156-761-3306

Alignment Scores:
 Pred. No.: 3,78e-28 Length: 3417
 Score: 346.50 Matches: 137
 Percent Similarity: 43.81% Conservative: 54
 Best Local Similarity: 31.42% Mismatches: 177
 Query Match: 15.42% Indels: 70
 DB: 14 Gaps: 15

US-09-985-689A-1 (1-434) x US-10-156-761-3306 (1-3417)

QY 18 GlyLeuTyrGlyGlnGlyIleValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
 Db 745 GGGTACGACGCGAAGGGCGTCAAGATCGCGCTTGGACACCGGTGTCGAC----- 795

QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
 Db 796 -----CGCACCCAGCGAGTCAAGACCGAGTGGCGGAGTCCCAAGAACTTCTCC 846
 QY 58 ArgThrAsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
 Db 847 GCGCGCGCGCGCGCGCGCGCGCTTCGGTTCAGCGCAGCGACGTCGCTCCATCGCGGCG 906
 QY 78 GlyAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 93
 Db 907 GGCACCGCGCGCGCGCGCGCGCGTACAGGAGTACAGGAGTGGCGCGCGCGCGCGCGCTC 966
 QY 94 PheGlnSerIleMetAspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThr 113
 Db 967 AACGGCAAGTCTCCGACGACACCGGC-----TCGCGCGACGACTCTCCGCTCCTGGCC 1020
 QY 114 LeuPheSerGlnAlaTyrSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAl 133
 Db 1021 GGCATGGAGTGGCGCGCGCGCGCGCGCTCGTCAACTGAGCTTGGCGCGCGCGCG 1080
 QY 133 aValAsnGlyAlaTyrThrAspSerArgAsnValAspAspTyrValArgLysAsnAs 153
 Db 1081 GACACCCCGAGATCGACCCGCTGGAAGCGGAG-GTCAACAAGCTCTCCGAGGAGAGGG 1139
 QY 153 pMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGly---GlyThrIleSerAl 172
 Db 1140 CATCCTCTTCGCGATCGCGCGCGCGCGCGAGTTCGGCGGAGCAGACCATCGGCTC 1199
 QY 172 aProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPh 192
 Db 1200 CCGCGCGCGCGCGCGCGCGCTCACGCTCGCGCGC----- 1236
 QY 192 eGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLy 212
 Db 1237 -----GTGAACGACAGCGACAGCTGGCTCTCTCCAGCGCGCGCGCGCT 1286
 QY 212 sAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSe 232
 Db 1287 GGACGGCGCCATCAAGCGCGCGCTCACCGCGCGCGGTGGAGATCATCCGCGCGC----- 1341
 QY 232 rSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLys-----TyrAl 249
 Db 1342 ----GCGCGCGCGCGCGCGCTACGACAGGAGTGGCGCGCGCGCGCGCTACCT 1397
 QY 249 aTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuAr 269
 Db 1398 CACCATCTCCGCTACGTGATGCGGACCGCGCGCTGTCGCGCGCGCGCGCGCTCTCAA 1457
 QY 269 gGluHisPheValLysAsnArgGlyIleThrProLysProSerLeu-----LeuLysAl 287
 Db 1458 GCAGCAGCAC-----CCCAACTGCTCGTTCGCGCGCGCTCAAGGG 1496
 QY 287 aAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTr 307
 Db 1497 CGCGCTGACCGCGCTCCGCGAAG-----GCGCGCAAGTACACGCGCTTCAGCAGGCGTC 1550
 QY 307 pGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerLe 327
 Db 1551 GGGCGGTATCGCGCTCGACAGCGATCAAGCAGTCCGCTATCGCCACCGCGACTCGGT 1610
 QY 327 uSerThrSerGlnLysAlaThrTyrSerPhe-ThrAlaThrAlaGlyLysProLeuLysI 347
 Db 1611 G-----AGCTTCGCGCATCCAGCAGTGGCGCGCGCGCGCGCGCGCG 1646
 QY 347 lSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuValAsnA 367
 Db 1647 CAAGCGCGGTCAACCGCAGCTACCTACCGCAACCTCGCGCAGCGTACGCTC-ACGCTGA 1705
 QY 367 sLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPhe---- 385
 Db 1706 ACCTCGCGTCGACGGCGCGCGCGCGCGCGCTCGCGCGCTCGCGCTCTCTCAACG 1765

QY 386 -----ThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValG 400
 Db 1766 TCGGCGCAGGAAGGTCACGGTCCCGCGGGC-----GGCAAGCCTCGGTGC 1813
 QY 400 lu-----AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValG 417
 Db 1814 ACTTCAGGTCAACAGGAAGCTGGCGGCACCGACCGCGGTACTCGCGGTACGTGA 1873
 QY 417 InAlaTyrAsnValProValGlyProGlnThrPheSerLeuAla 431
 Db 1874 CGGCCACGGC-----GGCGCCACACCGTCCGACGGCG 1908

RESULT 7

US-10-156-761-1/c

; Sequence 1, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIKA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; PRIOR FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 1
 ; LENGTH: 9025608
 ; TYPE: DNA
 ; ORGANISM: Streptomyces avermitilis
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (4187715)
 ; OTHER INFORMATION: a, t, c, g, other or unknown
 US-10-156-761-1

Alignment Scores:

Pred. No.: 5,11e-23 Length: 9025608
 Score: 346.50 Matches: 137
 Percent Similarity: 43.81% Conservatives: 54
 Best Local Similarity: 31.42% Mismatches: 177
 Query Match: 15.42% Indels: 70
 DB: 14 Gaps: 15

US-09-985-689a-1 (1-434) x US-10-156-761-1 (1-9025608)

QY 18 GlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
 Db 4132505 GGTACGACGGCAAGGCGCTCAAGATCGCGTCTCGACACCGCGGTGTCGAC----- 4132455
 QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
 Db 4132454 -----CGGACCCACCGGACCTCAAGGACCGAGTGGCGGAGTCCCAAGAACTTCTCC 4132404
 QY 58 ArgThrAsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
 Db 4132403 GCGCGCGCGACCGCGCGACCACTTCGGTTCACGGCAGCAGCTCGGTCCATCCGCGCG 4132344
 QY 78 GlyAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 93
 Db 4132343 GGCACGGCGCGCAAGTCCCAACGCAAGTACAGGGTGTCCGCGCGCGCGACGATCCTC 4132284
 QY 94 PheGlnSerIleMetAspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThr 113
 Db 4132283 AACGGCAGGTCTCTCGACACACCGGC-----TCCGGGCGACACTCCCGCATCTTGCC 4132230

QY 114 LeuPheSerGlnAlaTyrSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAl 133
 Db 4132229 GGCATGGAGTGGCGCGCGCAGCAGGCGCGCGTCTCACTCACTGAGCTGGCGCGGC 4132170
 QY 133 aValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArgIlyAsnAs 153
 Db 4132169 GACACCCCGGAGATCGACCCCGCTGGAAGCGAG-GTCAACACAGCTCTCCGAGGAGAGGG 4132111
 QY 153 pMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGly---GlyThrIleSerAl 172
 Db 4132110 CATCTCTTCGCGATCGCGCGCAGCAAGAGCGGAGTTCGGCGGACACACCATCGGCTC 4132051
 QY 172 aProGlyThrAlaIlyAsnAlaIleThrValGlyAlaThrGluAlaLeuArgProSerPh 192
 Db 4132050 CCCGGGAGCGCGCGCGCGCTCACCGCTCACCGTCCGCGC----- 4132014
 QY 192 eGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrIly 212
 Db 4132013 -----GTGAACGACGCGCAGCAAGCTGGCTCTCTCCAGCGCGCGCGCT 4131964
 QY 212 sAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSe 232
 Db 4131963 GGACGGCGCCATCAAGCCCGACGTCAACCGCACCCTGGGTGGACATCACCGCGCGC----- 4131909
 QY 232 rSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLys-----TyrAl 249
 Db 4131908 -----GCCGCGCGCGCGCTCATCGCACGAGGTCGCGCAGCGCGCGCGGTACCT 4131853
 QY 249 aTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuAr 269
 Db 4131852 CACCATCTCCGTTACGTGTCGACCGCGCATCTCGCGCGCGCGCGCATCTCTCAA 4131793
 QY 269 gGluHisPheValIlyAsnArgIlyIleThrProLysProSerLeu-----LeuLysAl 287
 Db 4131792 GCAGCAGCAC-----CCCACTGGTCTGTCGCGAGCTCAAGG 4131754
 QY 287 aAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTr 307
 Db 4131753 CGCGCTGACCGGCTCCCGAAG-----GGCGCAAGTACACGCGGTTCACAGGGGTC 4131700
 QY 307 pGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerLe 327
 Db 4131699 GGGCGGTATCGCGTCGACAAAGCGATCAAGCAGTCCGTGATCGCAACCGCGACCTCGT 4131640
 QY 327 uSerThrSerGlnIlyAlaThrTyrSerPhe-ThrAlaThrAlaGlyLysProLeuLysI 347
 Db 4131639 G-----AGCTTCGGCATCCAGCAGTGGCGCGCACACGACGA 4131604
 QY 347 leSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuValAsnA 367
 Db 4131603 CAAGCGCGTCACCCACGAGTCACTACCCACCTCGGACGAGTACGTC-ACGCTGA 4131545
 QY 367 sPleuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValIlyAsnAspPhe---- 385
 Db 4131544 ACTTCGCGTCGACGGCGGACCAACCCCAAGGGCTCGCGCGCTCTCTTCAAGC 4131485
 QY 386 -----ThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValG 400
 Db 4131484 TCGGCGCGCAGGAAGTCACTCCCGCGCGG-----GGCAAGCGCTCGGTGCG 4131437
 QY 400 lu-----AsnValPheIleAsnAlaProGlnSerGlyThrThrIleGluValG 417
 Db 4131436 ACTTCACGTCACCAACGAAGTGTGGCGGACCAACCGGACCGCGGTACTCGCGGTACGTCGA 4131377
 QY 417 InAlaTyrAsnValProValGlyProGlnThrPheSerLeuAla 431
 Db 4131376 CGGCCACGGC-----GGCGCGCACACCGTGGCGACGGCG 4131342

RESULT 8

US-09-974-300-1934
 ; Sequence 1934, Application US/09974300
 ; Patent No. US20020146721A1
 ; GENERAL INFORMATION:

QY	197	AspAsnIleAsnHisValalaGlnPheSerSerArgGlyProThrLysAspGlyArgIle	216
Dd	979	GAGCAT-----GTCCCTCTTATTCAAGCAGAGGGCGACAATCTATGGTCAAGTC	1029
QY	217	LysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSer-----SerLeu	234
Dd	1030	AAACCGGACTTGCTGGTACC GGGCACAAATATTACGTCGCTTCGTTCAACCGGATCTTTT	1089
QY	235	AlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrrAlaTyrMetGlyGlyThr	254
Dd	1090	CTCGATRAGCTGCCAAAAACAACACAGTCGGCACAAAATAATATGACATTGTC CGGAAC	1149
QY	255	SerMetAlaThrProIleValalaGlyAsnValalaGlnLeuArgGluHisPheVallys	274
Dd	1150	TGGATGGCTAGCCGATCTGCCGCAAAATTGCCGCA-----CTTATCCTTT	1194
QY	275	AsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAla	294
Dd	1195	CAGCAAGCCCCGGGCACAGACCTGATGAAGTCAAACAGCTGCTAATGGACGT---ACC	1251
QY	295	AspIle-----GlyLeuGlyTyrProAsnGlyAsnGln	305
Dd	1252	GATTTATGAAAGATCGCGATCCAAATGTTTTACGGTGCAGGTTACATCAACGCAGAACA	1311
RESULT 9			
US-10-090-624-5			
; Sequence 5, Application US/10090624			
; Publication No. US2002012335A1			
; GENERAL INFORMATION:			
; APPLICANT: TAKAKURA, Hikaru			
; APPLICANT: MORISHITA, Mio			
; APPLICANT: SHIMOJO, Tomoko			
; APPLICANT: ASADA, Kiyozo			
; APPLICANT: KATO, Ikunoshin			
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERHYMOSTABLE			
; FILE REFERENCE: TAKAKURA=6			
; CURRENT APPLICATION NUMBER: US/10/090.624			
; PRIOR FILING DATE: 2002-03-06			
; PRIOR APPLICATION NUMBER: 09/445,472			
; PRIOR FILING DATE: 1999-12-06			
; PRIOR APPLICATION NUMBER: 151969/1997			
; PRIOR FILING DATE: 1997-06-10			
; NUMBER OF SEQ ID NOS: 33			
; SOFTWARE: PatentIn version 3.0			
; SEQ ID NO 5			
; LENGTH: 4765			
; TYPE: DNA			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: Synthetic			
US-10-090-624-5			
Alignment Scores:			
Pred. No.:	1,6e-23	Length:	4765
Score:	307.00	Matches:	142
Percent Similarity:	38.92%	Conservative:	60
Best Local Similarity:	27.36%	Mismatches:	178
Query Match:	13.66%	Indels:	139
DB:	13	Gaps:	21
US-09-985-689A-1 (1-434) x US-10-090-624-5 (1-4765)			
QY	21	GlyGlNglGlyGlnIleValalavalAlaAlaSerThrGlyLeuAspThrGlyArgAsnAspSer	40
Dd	1024	GGCAATGGTTATGACATTGCATATGTGCATATGTGCATGCTGTTTACCTACTACTACGCT	108
QY	41	-----SerMetHisGluAlaPheArgGly	48
Dd	1084	GTCCCACTTGCCCAAGTACAACGTTACTTATGATGTTGCTGTTTACCTACTACTACGCT	114
QY	49	LysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAlaAsn-----	63
Dd	1144	CCYTCTCAACTACGCTGTTGCGAATAATAGTCTTCAACGAGAATATGCAGTATTCGCGTGG	120

```

; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1934
; LENGTH: 1329
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
; US-09-974-300-1934

Alignment Scores:
Pred. No.: 3,35e-27 Length: 1329
Score: 332.50 Matches: 104
Percent Similarity: 43.53% Conservative: 44
Best Local Similarity: 30.59% Mismatches: 109
Query Match: 14.80% Indels: 83
DB: 10 Gaps: 13

US-09-985-689A-1 (1-434) x US-09-974-300-1934 (1-1329)
QY 11 AspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAasp 30
    ::::: ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 406 GAAGTGTCTGAGAACAAATCAGACGCTGACAGCGCAAGAGCATGCGTGTCTGTCATTGAT 465
    ::::: ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 31 ThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIle 50
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 466 ACGGGCGTA-----TACGCTCACGAAGATCTTGAAGGCAGGATC 504
    ||||| ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 51 ThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsnGlyHisGly 69
    ||||| ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 505 AGGCGTTTCAAGACTTTATCAACCAGACAGACACACCTATGATGACATGGCACGGC 564
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 70 ThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsn-----LysGly 85
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 565 ACACACTGCGCGGTGATGCTTGGCGAAACGAGGCGGCTCATCGGGTACGTACCGCGGA 624
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 86 MetaLproGlnAlaAsnLeuValPheGlnSerIleMetAspSerGlyGlyGlyLeuGly 105
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 625 CCTGCTCTGAAGCAGAACTTGTGCGGTGTAAGATTTGGACAAATGGGATCC---GGA 681
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 106 GlyLeuProSerAsnLeuGln-----TGGTGGTGCATTCATTAATCAATGAAGAAATCCT 112
    ||| ||| ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 682 TCGCTCAAAACCGTCATCAAGCGCTAGATTGGTGCATTCATTAATCAATGAAGAAATCCT 741
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 113 -----ThrLeuPheSerClnAlaTyrSerAlaGlyAlaArgIleHisThrAsn 128
    ::::: ||| ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 742 GATGATCGATCGACATATTTCATGTGTCATGGTGCGAAGCCCTTCGCTACGAGAT 801
    ::::: ||| ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 129 -----SerTrpGlyAlaAlaValAsnGly 136
    ::::: ||| ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 802 GAAGNAGAAGATCCAGTCGTTAAAGCTGTTTCATGCAGCATGGACGCGCAGGCATT----- 855
    ::::: ||| ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 137 AlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIle 156
    ::::: ||| ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 856 -----GTTGTA 861
    ::::: ||| ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 157 LeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAla 176
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 862 TGTGCGGCGAGCGGCAACTCCGCTCTGATGGCAACAGATTGCCAGCCCGGGTGTCTCAGC 921
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 177 LysAsnAlaIleThrValIcIyAlaThrGluAsnLeuArgProSerPheGlySerTyrAla 196
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 922 AGCAAGATTATTACACTCGGACCCCTGGATGAC---AGGGATACAGTCAAGTCAGCCGAGGAT 978
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Oy 64 AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsn 83
 Db 1204 GATGTCAGGGTCACGGAACCTCACGTAGCTGGAACTGTTGCTGTTACGACAGCAACAAT 1263
 Oy 83 ----- 83
 Db 1264 GATGCTTGGGATGGCTCAGTATGCTACTCTGTTGAATGGGAAGTGTCTCAAGACTCTAT 1323
 Oy 84 -----LysGlyMetAlaProGlnAlaAsn 91
 Db 1324 GGTGGGATATACGAACGTTTACCACAGACACCGTGCAGGGTGTCTCCAGGTGCCCA 1383
 Oy 92 LeuValPheGlnSerIleMetAspSerGlyGlyLeuGlyGlyLeuProSerAsnLeu 111
 Db 1384 ATAATGGCAATAAGAGTCTTCTAGGAGT---GATGGACGGGTAGCATGTGGGATATTATA 1440
 Oy 112 GlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGly 131
 Db 1441 GAAGGTATG---ACATACGCACCAACCCATGTCGACAGCTTATAAGCATGAGTCTCGCT 1497
 Oy 132 AlaAlaValAsnGlyAlaTyrThr-----ThrAspSerArgAsn-----ValAspAsp 147
 Db 1498 GCA-----AATGCTCCACTACTAGATGCTACTGATCCAGAAAGCGTGTCTGTGGATGAG 1551
 Oy 148 TyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGly 167
 Db 1552 CTTACCGAAAAGTACGGTGTCTGTTATTCGTAATAGCTGCAGGAAATGAAGTCTCGCAT 1611
 Oy 168 GlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsn 187
 Db 1612 AACATCGTGGAGTCTCGGTGTTGCAACAAAGGCAATAAATGTTGGAGCTGCTGCA--- 1668
 Oy 188 LeuArgProSerPheGlySerTyrAlaAsp----- 197
 Db 1669 GTGCCCAATTAAAGTGTGGAGTTATGTTTCCCAAGCACTTGGATATCTCTGATTACTATGGA 1728
 Oy 198 -----AsnIleAsnHisValAlaGlnPheSerSerArgGly 209
 Db 1729 TTCTATTACTTCCCGCCTACACAAAGCTT---AGAATAGCATCTTCTCAAGCAGAGGG 1785
 Oy 210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSer 229
 Db 1786 CCGAGAATAGATGTTGAATAAACCACCAATGATGCTCCAGGTTCAGGAATTTATACPCA 1845
 Oy 230 AlaArgSerSerLeuAlaProAspSerSerPheThrAlaAsnHisAspSerLysTyrAla 249
 Db 1846 TCCTCGCGATGTGGATTGGCGGAGCTGACTTC----- 1878
 Oy 250 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArg 269
 Db 1879 ---ATGCTGGAACTTCGATGGCTACTCCACATGTCAGCGGTGTGCTGCACTCTCATA 1935
 Oy 270 GluHisPheValLysAsnArgGlyIleThrProLysProSerLeuLysAlaAlaLeu 289
 Db 1936 AGCGGG---GCAAAGCGCAGGGAATATATACTAATCCAGATATATAATTAAGAAGTCTT 1992
 Oy 290 IleAlaGlyAlaAlaAspIle-----GlyLeuGlyTyrProAsnGly 303
 Db 1993 GAGACGGGTGCACCTCGGCTTGAGGAGATCCATATATCTAGGCGGATGCTGCTGCACTGAGCTT 2052
 Oy 304 AsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlu 323
 Db 2053 GACCACAGTCTCCCAATTTGTAACGTTACCAAGTCTCTGGGAAATCCTTAAGGCTATATAAC 2112
 Oy 324 SerSerLeuSerThrSerGln-LysAlaThrTyrSerPheThrAlaThrAlaGlyLy 343
 Db 2113 GGCACCACTCTCCCAATTTGTAACGTTACCAAGTCTCTGGGAAATCCTTAAGGCTATATAAC 2164
 Oy 343 sProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAlaSerValTh 363
 Db 2165 ---TTGCGGAGTACTTGGGTGGACGTTATAAGAGGCTCTCTACCAAGGAACCT----- 2215

Oy 363 rLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThr--GlnTyrValGly 382
 Db 2216 ---CTATACCTGACATTCGCGATGCGGACACATTAAAGTACGTAGG 2256
 Oy 383 Asn-----AspPheThrSerProTyrAsnAspAsnTrp----- 393
 Db 2257 GACACGGAGTACAGAACTTTTGGATCTATGCAACTGAGCCATGGATTAAAGCCTTTTGTG 2316
 Oy 394 AspGly-----ArgAsnAsnValGluAsnValPhe-----Ile 404
 Db 2317 AGTGAAGTGTAAATCTTAGAACAATAACCGAGTTTGTCTTAGGTTGAATATGATGTA 2376
 Oy 405 AsnAlaProGlnSerGlyThrTyr----- 412
 Db 2377 GAGGTCCTTGAGCCAGCTCTCTATGTTGGAAGGATAATCATCATGATCATCAACAAGCCCA 2436
 Oy 413 ThrIleGluValGlnAlaTyrAsnValProValGlyProGlnThrPheSer 429
 Db 2437 GTTATTGAAGACGAGATCTTGAACACAATTTGTTATTCGCGAGAGTTCACT 2487

RESULT 10

US-09-927-827-29
 ; Sequence 29, Application US/09927827
 ; Publication No. US20030036176A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bower, Stanley G.
 ; APPLICANT: Ramseier, Thomas M.
 ; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
 ; FILE REFERENCE: 38-10(15824)B
 ; CURRENT APPLICATION NUMBER: US/09/927,827
 ; CURRENT FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: US 60/279,493
 ; PRIOR FILING DATE: 2001-03-28
 ; NUMBER OF SEQ ID NOS: 69
 ; SEQ ID NO 29
 ; LENGTH: 3743
 ; TYPE: DNA
 ; ORGANISM: Xanthomonas campestris
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1001)..(2743)
 ; US-09-927-827-29

Alignment Scores:

Pred. No.: 5,34e-21 Length: 3743
 Score: 283.00 Matches: 127
 Percent Similarity: 37.53% Conservative: 55
 Best Local Similarity: 26.19% Mismatches: 147
 Query Match: 12.59% Indels: 156
 DB: 11 Gaps: 24

US-09-985-689A-1 (1-434) x US-09-927-827-29 (1-3743)

Oy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeu-----AspThrGly 36
 Db 1502 GCGACGGCGACCGCTCGTGGCGTGTATGATACCGGCATCACCAGTCATGCGGACCTCAAC 1561
 Oy 37 ArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeu 56
 Db 1562 GCCAACATCTCTCGCGGGTACGACTTCATCAGCGATGCGACCCACCGCGCATGGCAAC 1621
 Oy 57 GlyArgThrAsnAsnAlaAsnAsp----- 64
 Db 1622 GCGCGTGACAGCAACGCCGCCACGAGGCGACTGTAGCCGCCAACGAATGCGGCGCC 1681
 Oy 65 -----ThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGly 78
 Db 1682 GGCATTCCCGCGCCAGCTCCAGCTGGCAGCCACCCATGTGCGCCGCGCAGGTTCGCGGCA 1741
 Oy 79 AsnGlySerThrAsnLysGlyMetAlaProGlnAla----- 90
 Db 1742 GTGACCAACAACACCACCGCGGTAGCGCGCACCGCTACGGGCGCAAGGTGCTGACCGGTG 1801

QY	91	-----AsnLeuValPheGlnSerIleMetAsp	99	Db	2663	ACCGCACC	-----TCGGAACGTATTACGTGCGTCTGAAGCGCTACAGC	2707
Db	1802	CGCGTGTCCGGCAAGTCCGCTGGTGGTCTGCTGCAGATATCCGCCGACGCATCGTCTGGGCC	1861	QY	425	ProGlnThrPheSer	429	
QY	100	SerGlyGlyLeuGlyLeuProSerAsnLeuGln	117	Db	2708	-----ACGTTCTCC	2716	
Db	1862	TCCGGGGCACCGTACGCGCATCCCGCCCAATCTAAACCGCGCGAGGTGATCAACATG	1921					
QY	118	AlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAla	137					
Db	1922	TCGCTCGCGCGCGGTAGTCTGCACACCACCATGCAGAAC	1978					
QY	138	TyrThrThrAspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLeu	157					
Db	1979	GTGTCCGCGCGCAC	2002					
QY	158	PheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLys	177					
Db	2003	GTCGCGCGCGCACGATGCGTCAATGTCCGGT	2059					
QY	178	AsnAlaIleThrValGlyAlaThrGluAsn	195					
Db	2060	AACGTGATTGGGTGGCGCGCACCATCTCGCGCGCGGAGGCGCAGTATTCCAACTTC	2119					
QY	196	AlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArg	215					
Db	2120	GGCACCGGTATC	2131					
QY	216	IleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSer	233					
Db	2132	-----GATGTGTCCGCGCGCGGTCTGCTGCATCTCTCCACCTCAACAGCGCGCAC	2182					
QY	234	LeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrValatyrMetGlyGly	253					
Db	2183	ACCAGCGCGGGTAGC	2221					
QY	254	ThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheVal	273					
Db	2222	ACCTCGATGGGTCCCGCATGTGGCGCGGTGGTGGCGTGGTGCAGTCG	2275					
QY	274	LysAsnArgGlyIleThrProLys	289					
Db	2276	GCCCGCAGCGCGCTGACGCCACCGCGGTGGAACTTGTGAAGAACACCGCGCGTGTCT	2335					
QY	290	IleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArg	309					
Db	2336	TTACCGCGCGCGCTGC	2374					
QY	310	ValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSer	325					
Db	2375	GTCACGCGCATGCCGCGTCACTGCGGCC	2431					
QY	326	-----SerLeuSerThrSer	330					
Db	2432	GGTGTGGAAACACCTTGACCAAGGCACCTCCGGTGCACCGCGCTGGCGCGGCGACTGGC	2491					
QY	331	GlnLysAlaThrTyrSerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuVal	350					
Db	2492	GCGGAATGAATACACCATCACCGTCCGCGCGCGC	2527					
QY	351	TrpSerAspAlaProAlaSerThrThrAlaSerValThrLeu	367					
Db	2528	-----AGCGGCACCTTGACCGTGCACCAACACCGCGCGCGCGCGCAT	2569					
QY	368	LeuAspLeuValIle	384					
Db	2570	GCGGACCTGTATGTGGCGCGCGCATGTGCACCGACCTGCGGCTTACACC	2620					
QY	385	PheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGluAsnValPheIle	404					
Db	2621	-----TGCGCGCCCATACCGCAC	2662					
QY	405	AsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGly	424					

```
Db 2201 GTCAATCTAGCTGGCGGGTACTTCGATCCGAAAGACTACGGCTGTGGCTTC----- 2254
QY 137 AlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIle 156
Db 2255 -----ACGCGGTATATCAACAGAGTTCGGCGGGTTATGGCGGAGGGCGTACTGGTA 2305
QY 157 LeuPheAlaAlaGlyAsnGluGly-----ProAsnGlyGlyThr--- 169
Db 2306 GTGGTGGCGCGCGCAACGAAGCGCTGGCTGGCTGATGCAGACGACGGCGGCACCTAT 2365
QY 170 -----IleSerAlaProGlyThrAlaLysAsnAlaIleThrVal 182
Db 2366 CCGCGCAACATGATCTGTCAGCAGCGATCCGGGCAATCTGGAGGACGCGATCGTGGTG 2425
QY 183 GlyAlaThrGluAsnLeuArgPro---SerPheGlySerTyrAlaAspAsnIleAsnHis 201
Db 2426 GGATCGGTGCACAGACGACCGCCGACAAATACGGC----- 2461
QY 202 ValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspValMet 221
Db 2462 GTGTCTATATTTTCATCCGCGGGCGGACCGCGATGGCGGCTCCAAACCTGATGTGTC 2521
QY 222 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 241
Db 2522 GCGCGCGCGGAAAGATCTCTCGCTTACTACGGCTTCGACCGC----- 2566
QY 242 AlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 261
Db 2567 CGCGACCCGTCACGCTGATGTCGAGATCAGCGGCGCACCATGTCGCGCGCACCGCATGTG 2626
QY 262 AlaGly-----Asn 264
Db 2627 TCCGGCGTCTGCGCGGGTTTTATCCGACGCGCGAGTTCATCGGCTTTCGACCGG 2686
QY 265 ValAlaGlnLeu-----ArgGluHisPheValLysAsn 275
Db 2687 GTCAAGCAACTGCTGCTGCACACCTGCACCGACCTGCAGCGCGATCGTTACGTGCGAGGGC 2746
QY 276 ArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAsp 295
Db 2747 AGGGGGGTG-----CCGAATTGTATCGCG-----ATGCTTGGAGAGACGTGA 2788
QY 296 IleGly-----LeuGlyTyrProAsnGlyAsnGlnGly 306
Db 2789 TTCGGGAGTGGGATTTGGGATTCGCAACACGCGGTGGTGGC 2830
```

RESULT 12

```
US-09-966-921A-1
; Sequence 1, Application US/09966921A
; Patent No. US20020103100A1
; GENERAL INFORMATION:
; APPLICANT: Outtrup, Helle
; APPLICANT: Lassen, Soren
; APPLICANT: Pedersen, Poul
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides Having Proteolytic Activity
; FILE REFERENCE: 10097.200-US
; CURRENT APPLICATION NUMBER: US/09/966,921A
; CURRENT FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1306
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (50)..(1303)
; OTHER INFORMATION:
; NAME/KEY: mat_peptide
; LOCATION: (371)..()
; OTHER INFORMATION:
; US-09-966-921A-1
```

```
Alignment Scores:
Pred. No.: 6,66e-21 Length: 1306
Score: 276.00 Matches: 100
Percent Similarity: 44.59% Conservative: 40
Best Local Similarity: 31.85% Mismatches: 110
Query Match: 12.28% Indels: 64
DB: 10 Gaps: 17

US-09-985-689A-1 (1-434) x US-09-966-921A-1 (1-1306)

QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 446 GCGAGCGGAATTAAGTAGTACTGTTTAGATACAGGGGTTTATACAGC----- 493
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 494 -----CATTTAGATTTAGCTGCTCTCGCGAGCAATGCAAGGATTTTACCAATCTAAT 547
QY 61 -----AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySer 75
Db 548 CCTTTAGTAGATGGTTTCATGCCCGCATGCCAAGGCGATGTTACATATGTCGCGAAT 607
QY 76 ValLeuGlyAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsn 91
Db 608 GTATTGGCGCATGGAGCGCAATAATGCACAAGGCGTTTACGGGTGGCTCCGCAAGCGAAA 667
QY 92 Leu---ValPheGlnSerIleMetAspSerGlyGlyLeuGly---GlyLeuProSer 109
Db 668 CTATGGGCATATAAAGTATTAGGAGATAACGCGCGGATCTCTGATGATATTGCGAGCA 727
QY 110 AsnLeuGlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsn--- 128
Db 728 GCTATCACACATGTAGCTGATGAGCTTCACGTACAGGTTCCAAAGTAGTAATTAATATG 787
QY 129 SerTrpGlyAlaValAlaAsnGlyAlaTyrThrAspSerArgAsnValAspAspTyr 148
Db 788 TCGTAGGTTCATCTGCCAAGGATTCATTGATTGCT-----AGTCAGTAGATTAT 838
QY 149 ValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGly 168
Db 839 GCATATGGAAGAGGTGTTAATCTGCTGCGGTGTTAATAGTGGTCCAGGCGACCAAT 898
QY 169 ThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeu 188
Db 899 ACAATCGGCTTTCCTGGCGGGCTTCTAATGAGTCGCGGTGTTAATAGTGGTCCAGGCG 958
QY 189 ArgProSerPheGlySerTyrAlaAlaAspAsnIleAsnHisValAlaGlnPheSerArg 208
Db 959 CAGCAAAAT---GGAACCTAT-----CGAGTAGCTGATTCTCATCTAGTA 1000
QY 209 Gly---ProThrLysAspGly-----ArgIleLysProAspValMetAla 222
Db 1001 GGGATCCGCGCACTGCTGGAGATTATATCATTCAGAGCGGTGATATTCAAGTTTCAGCT 1060
QY 223 ProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPheTrpAla 242
Db 1061 CCGGAGCAAGTGTA-----GAGTCTACATGG--- 1087
QY 243 AsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAla 262
Db 1088 ---TACACTGGCGGTTAATATACGATCAGCGGTACATCAATGGCTACACCTCATGTAGCT 1144
QY 263 GlyAsnValAla-----GlnLeuArgGlu 270
Db 1145 GGGTAGCTGCTAAATCTGTCAGCGAATACTTCATTAAAGTCATAGCACTGCGGCACA 1204
QY 271 HisPheValLysAsnArgGlyIleThrProLysProSerLeuLysAlaAlaLeuIle 290
Db 1205 GAA---TTGCAAAATCGCGCT-----AAAGTATATGATATTAAAGGTGGTATCGGA 1252
QY 291 AlaGlyAlaAlaAsp-----IleGlyLeuGlyTyrPro 301
Db 1252 ----- 1306
```

Db	1253	GC	CGGAACAGCTGACGATTATGATCATCAGGTTCCGATATCCA	1294
RESULT 13				
US-09-966-921A-5				
; Sequence 5, Application US/09966921A				
; Patent No. US20020103100A1				
; GENERAL INFORMATION:				
; APPLICANT: Outtrup, Helle				
; APPLICANT: Lassen, Soren				
; APPLICANT: Pedersen, Poul				
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides Having Proteolytic Activity				
; FILE REFERENCE: 10097 200-US				
; CURRENT APPLICATION NUMBER: US/09/966,921A				
; CURRENT FILING DATE: 2001-09-28				
; NUMBER OF SEQ ID NOS: 5				
; SOFTWARE: PatentIn version 3.1				
; SEQ ID NO 5				
; LENGTH: 1330				
; TYPE: DNA				
; ORGANISM: Artificial Sequence				
; FEATURE:				
; OTHER INFORMATION: Synthetic				
US-09-966-921A-5				
Alignment Scores:				
Pred. No.:	6.85e-21	Length:	1330	
Score:	276.00	Matches:	100	
Percent Similarity:	44.59%	Conservative:	40	
Best Local Similarity:	31.85%	Mismatches:	110	
Query Match:	12.28%	Indels:	64	
DB:	10	Gaps:	17	
US-09-985-689A-1 (1-434) x US-09-966-921A-5 (1-1330)				
Qy	21	GlyClnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer	40	
Db	406	GGCAGCGGAATTAAAGGTAGCTGTTTATAGATACAGGGGTTTATACAAAGC-----	453	
Qy	41	SerMethHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn	60	
Db	454	-----CATTTAGATTAGCTGGTCTGCCAGCAATGCAAGGATTTTACCAATCTAAT	507	
Qy	61	-----ASNAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySer	75	
Db	508	CTTTAGTAGATGCTTCATGCACCGATCGCCAAAGGCGATGTACACATGTTCGCGAAT	567	
Qy	76	ValLeuGlyAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsn	91	
Db	568	GTATTGGCGCATGGAGCGAGTAATGGCAAGGCGTTTACGGGGTGCTCCGCAAGCGAA	627	
Qy	92	Leu---ValPheGlnSerIleMetAspSerGlyGlyGlyLeuGly---GlyLeuProSer	109	
Db	628	CTATGGGCATATAAGATTAGGAGATAACGCGCGGATCTCTGATGATATTGCAGCA	687	
Qy	110	AsnLeuGlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsn---	128	
Db	688	GCTATCAGACATGTAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG	747	
Qy	129	SerTrpGlyAlaAlaValAsnGlyAlaTyrThrAspSerArgAsnValAspAspTyr	148	
Db	748	TCGTAGTTTCATCTGCCAAGGATTCATGATTGCT-----AGTCAGTAGATTAT	798	
Qy	149	ValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGly	168	
Db	799	GCATATGAAAGGTTGTTAATCTGTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	858	
Qy	169	ThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValClyAlaThrGluAsnLeu	188	
Db	859	ACAATCGCTTTCCTGGCGGCTTCTAAATGAGTGGCGAGTACGGCGATTTGGAGAATGT	918	
Qy	189	ArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArg	208	
Db	919	CAGCAAAAT---GGAACATTAT-----CGAGTAGCTGATTCTTCATCTAGA	960	
Qy	209	Gly---ProThrLysAspGly-----ArgIleLysProAspValMetAla	222	
Db	961	GGGAATCCGGCAACTGCTGGAGATTATATCATTAAGACGGTGATATTGAAGTTTCAGCT	1020	
Qy	223	ProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAla	242	
Db	1021	CCGGAGCAAGTGTA-----GAGTCTACATG---1047		
Qy	243	AsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAla	262	
Db	1048	---TACACTGCGGTTATAACGATCAGCATCAGCGGTACATCACTGCTACCTCATGTAGCT	1104	
Qy	263	GlyAsnValAla-----GlnLeuArgGlu	270	
Db	1105	GGTGTAGCTGCTAAATCTGTCAGCAATACTTCTAATTAAGTCTACGCCAAGTCCGCACA	1164	
Qy	271	HisPheValLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaLeuIle	290	
Db	1165	GAA---TTGCAAAATCGCGCT-----AAAGTATATCATATTAAGGTGTTATCGGA	1212	
Qy	291	AlaGlyAlaAlaAsp-----IleGlyLeuGlyTyrPro	301	
Db	1213	GCCGGAACAGGTGACGATTATGATCATCAGGGTTCGGATATCCA	1254	
RESULT 14				
US-10-156-761-5384				
; Sequence 5384, Application US/10156761				
; Publication No. US20030119018A1				
; GENERAL INFORMATION:				
; APPLICANT: OMURA, SATOSHI				
; APPLICANT: IKEDA, HARUO				
; APPLICANT: ISHIKAWA, JUN				
; APPLICANT: HORIKAWA, HIROSHI				
; APPLICANT: SHIBA, TADAYOSHI				
; APPLICANT: SAKAKI, YOSHIYUKI				
; APPLICANT: HATORI, MASAHIRA				
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES				
; FILE REFERENCE: 249-262				
; CURRENT APPLICATION NUMBER: US/10/156,761				
; CURRENT FILING DATE: 2002-05-29				
; PRIOR APPLICATION NUMBER: JP 2001-204089				
; PRIOR FILING DATE: 2001-05-30				
; PRIOR APPLICATION NUMBER: JP 2001-272697				
; PRIOR FILING DATE: 2001-08-02				
; NUMBER OF SEQ ID NOS: 15109				
; SEQ ID NO 5384				
; LENGTH: 3303				
; TYPE: DNA				
; ORGANISM: Streptomyces avermitilis				
; FEATURE:				
; NAME/KEY: CDS				
; LOCATION: (1)..(3303)				
US-10-156-761-5384				
Alignment Scores:				
Pred. No.:	1.25e-19	Length:	3303	
Score:	270.00	Matches:	112	
Percent Similarity:	37.61%	Conservative:	58	
Best Local Similarity:	24.78%	Mismatches:	168	
Query Match:	12.02%	Indels:	114	
DB:	14	Gaps:	17	
US-09-985-689A-1 (1-434) x US-10-156-761-5384 (1-3303)				
Qy	68	HisGlyThrHisValAlaGlySerValLeuGlyAsn-----GlySerThrAsnLys	84	
Db	1168	CACGCGACGCGCTCGCGGCATCAGCGCGCAACGCCCTGTTTCGCGCGCAAGATGAAC	1227	
Qy	85	GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGlyGlyGly---	103	
Db	1228	GGCGCGCGCGCGCGCGCGAAGATCGTCTCTCTCGCGTGCCTGACACTGACCGCGGCGTGC	1287	

QY 103 ----- 103
 Db 1288 ACCAAGTGGCCCTACCGAGGCGATGATCGACCTCGTCCCAACCGTGGCGTGACATC 1347
 QY 104 -----LeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyr 119
 Db 1348 GTCAACATGTCTATCGCGCTGCTCGCGG-----CTGAACAGCGGCAAC 1392
 QY 120 SerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 139
 Db 1393 AACGGCGCGCGACTCTACAG----- 1416
 QY 140 ThrAspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLeuPheAla 159
 Db 1417 -----CGTCTCATCGACACCTAC-----GGCGTCCAGCTGGTGATCTCC 1455
 QY 160 AlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla 179
 Db 1456 CGGGCAACTCGGGCCCGCGGCCAACACCATCGCGCGCCGCTGCGCCGCAACAAGTC 1515
 QY 180 IleThrValGlyAlaThr-----GluAsnLeuArgProSerPheGlySerTyrAla 196
 Db 1516 ATCTCGTGGCGCGCCATCTCCAAAGCAGACCTGGCGCGCCCACTACGGCTCCCGAGGTG 1575
 QY 197 AspAsnIleAsnHisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIle 216
 Db 1576 GAGAACGGCTACGGCATGATCGCTCTCTCGCGCGCGCGCTGAGGACGCGCGCTTC 1635
 QY 217 LysProaspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeu----- 234
 Db 1636 ACGGCGAGCTGACCGCGCCCGCGCCCTCGATCAACTCCACACGACCTGGCTGGCGGCG 1695
 QY 235 AlaProaspSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThr 254
 Db 1696 TCCCGGTGGCGAGGCGGTACTCTCGTGGCGCGCTACTCGATCTCGATCTCGAGGCGACG 1755
 QY 255 SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLys 274
 Db 1756 TCGATGGCTGCTCGCGAGCGCGCGCGCTTGGCGCGCTGCTGCTGAGC---GCCGCAAC 1812
 QY 275 AsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAla 294
 Db 1813 CACAAGGCATCGAGCTACCGCCCGCCACCTCGCGCGCGCTGCGCGCGCTCGACCGCGAC 1872
 QY 295 Aspile-----GlyLeuGlyTyrProAsnGlnGly 306
 Db 1873 CACATCAAGGTGTGTCAGCGGTACGAGGAGGCGCGCGCTCATCAACATCTGTGACGCG 1932
 QY 307 TrpGlyArgVal----- 310
 Db 1933 TGGAGTCCATCAGGAGGAGCGCCACGCGCGCACGACTACACCGTGAAGCGCGCGTGCAC 1992
 QY 311 ---ThrLeuAspLysSerLeu-----AsnValAlaTyrValAsnGluSer 324
 Db 1993 ACCGCGATCGACGCGGTGAAGACGCGCGGCTTGGCAGCGCGCTGTACGACCGCGAG 2052
 QY 325 SerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr----- 340
 Db 2053 GCGCGGCTCAAGCGCGGTGACAGAGACGTACGACGTACCATCAGCGGTGCGCGG 2112
 QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 Db 2113 GCCGACGCGGATCCCGGACGAGCTGTACTTCGAGAACACACGCGCGGTACCTCGCG 2172
 QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaPro----- 375
 Db 2173 ATCGTCGCG---TCGACGAGGTCAAGCTCCCGCTGAACGCGCGGTGACCGTCAAGTGC 2229
 QY 376 -----AsnGlyThrGlnTyrValGlyAsnAspPheThrSerProTyr 389
 Db 2230 CAGGCGCGCCCGAAGTGGCGCGCTGAAGAGCGGATCTCGAGGTCGACGACCGCGCG 2289
 QY 390 AsnAspAsnTrpAspGlyArgAsnValGluAsnValPheIleAsnAlaProGlnSer 409

Db 2290 ACCGAGGCGGTGCAC---AAGCAGATCTGTTCGACGCTGTCTCCGCGCGTCAAG 2346
 QY 410 GlyThrTyrThrIle-----GluValGlnAla-----TyrAsn 420
 Db 2347 TACAGTACTCGCGTCTCGGTTCGTCGAGCGCAACAGCAGCAGTCGTACTTCGTGTCAG 2406
 QY 421 ValProValGlyProGlnThrPheSerLeuAlaIle 432
 Db 2407 GTCCCGAGGCGCGCAAGTCCCTCGAGGTGCGGATC 2442

RESULT 15
 US-09-974-300-1938
 ; Sequence 1938, Application US/09974300
 ; Patent No. US20020146721A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Berka, Randy M.
 ; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
 ; TITLE OF INVENTION: Expression
 ; FILE REFERENCE: 10085-500-US
 ; CURRENT APPLICATION NUMBER: US/09/974,300
 ; PRIOR FILING DATE: 2001-10-05
 ; PRIOR APPLICATION NUMBER: 09/680,598
 ; PRIOR FILING DATE: 2000-10-06
 ; PRIOR APPLICATION NUMBER: 60/279,526
 ; PRIOR FILING DATE: 2001-03-27
 ; NUMBER OF SEQ ID NOS: 8481
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1938
 ; LENGTH: 1485
 ; TYPE: DNA
 ; ORGANISM: Bacillus licheniformis
 US-09-974-300-1938

Alignment Scores:
 Pred. No.: 1.38e-18 Length: 1485
 Score: 256.00 Matches: 94
 Percent Similarity: 42.82% Conservatives: 52
 Best Local Similarity: 27.57% Mismatches: 123
 Query Match: 11.39% Indels: 72
 DB: 10 Gaps: 16

US-09-985-689a-1 (1-434) x US-09-974-300-1938 (1-1485)

QY 11 AspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAsp 30
 Db 573 GATGTCTGGAAGAAAGCGCTAT-----AAAGGCAAGAGGCGCGTCTGCTCAATTGAC 626
 QY 31 ThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArgGly----- 48
 Db 627 ACCGGGTAGATGG-----GATCATCCGCGGTGAAACAAAGTACAGAGGTACAAAC 680
 QY 49 -----LysIleThrAlaLeuTyrAlaLeuGlyArgThr 59
 Db 681 CCTTCTAAACCGAACAGTCCAGACATGAGTTCACACTGGTTGATCGGTGAAGCAATAAG 740
 QY 60 AsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
 Db 741 AAAACGCCATATGATCAGTTGGGCCCGCACACATGTGACAGCAGCATGTCGCGTCA 800
 QY 80 -----GlySerThrAsnLysGlyMetAlaProGlnAlaAsn---LeuValPheGln 95
 Db 801 GAAGCGCGCGCAAAACACAGATCGGTGTGCGCGCGGCGGCAAAATGGATAGTGTCAAG 860
 QY 96 SerIleMetAspSerGlyGlyLeuGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPhe 115
 Db 861 GCCTTTTCAGAGACGCGGAGATGAAAAATCCCTTTTGGCTGCAGGAGATGATTTTG 920
 QY 116 SerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsn 135
 Db 921 GCCCGCAAGATGCGAAAGGA---AAAGCGCACCTGAAAGGCTCCCGATGTCGTCAAC 977

Search completed: July 28, 2003, 11:48:02
Job time : 10802.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 18:57:50 ; Search time 16.0185 Seconds
(without alignments)
2605.570 Million cell updates/sec

Title: US-09-985-689A-1
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAOSSYGLY.....EVQAYNVPVGPQTFSLAIVN 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_76: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	523.5	23.3	1743	2 T18279	multidrug resistan
2	497	22.1	1905	2 T18267	multidrug resistan
3	347.5	15.5	444	2 B83891	intracellular alka
4	320.5	14.3	442	2 A69587	intracellular alka
5	310.5	13.8	806	2 A41341	microbial serine p
6	302.5	13.5	1398	2 T28159	pyrolysin (EC 3.4.
7	283	12.6	580	2 S11890	serine proteinase
8	281	12.5	419	1 S25835	subtilisin (EC 3.4
9	280	12.5	799	2 G83753	subtilisin-type pr
10	279.5	12.4	1345	2 T29090	surface layer-asso
11	277	12.3	420	1 S23407	subtilisin (EC 3.4
12	267	11.9	715	2 JC4908	alkaline serine pr
13	263	11.7	1331	2 A72647	probable surface l
14	259.5	11.5	757	2 C84120	subtilisin-type pr
15	257.5	11.5	513	1 A35742	aqualysin (EC 3.4.
16	253.5	11.3	894	2 F69730	cell wall-associat
17	251	11.2	627	2 D75392	serine proteinase,
18	246.5	11.0	402	1 JU0332	alkaline proteinas
19	246.5	11.0	534	1 JS0173	alkaline proteinas
20	243	10.8	519	2 S71451	halolysin R4 (EC 3
21	242.5	10.8	401	2 I39978	serine proteinase
22	240	10.7	380	2 A49778	high-alkaline seri
23	240	10.7	488	2 A11930	protease (import
24	235.5	10.5	382	1 SUBSN	subtilisin (EC 3.4
25	235	10.5	378	2 A33973	high-alkaline seri
26	235	10.5	382	2 I39780	subtilisin (EC 3.4
27	234	10.4	910	2 C69456	subtilisin sendai
28	234	10.4	1374	2 D72593	hypothetical prote
29	232.5	10.3	525	2 G84406	halolysin (importe

30 231 10.3 321 1 S27501 alkaline proteinas
31 230.5 10.3 379 1 SUBSCL subtilisin (EC 3.4
32 229.5 10.2 601 2 J04576 serine proteinase
33 227 10.1 A35066 streptococcal C5a
34 226.5 10.1 1118 2 H97298 subtilisin like pr
35 225.5 10.0 381 2 JH0778 subtilisin (EC 3.4
36 225.5 10.0 613 2 S75976 hypothetical prote
37 225.5 10.0 1052 2 T17093 intraluminal subti
38 223.5 9.9 381 1 SUBSI subtilisin (EC 3.4
39 223.5 9.9 381 1 SUBSI subtilisin (EC 3.4
40 223.5 9.9 381 2 J01487 subtilisin (EC 3.4
41 222.5 9.9 275 2 J01085 subtilisin (EC 3.4
42 222 9.9 384 2 J04802 alkaline proteinas
43 221.5 9.9 272 2 A23624 subtilisin (EC 3.4
44 221.5 9.9 530 2 A42605 halolysin (EC 3.4
45 221 9.8 1036 2 J05568 serine proteinase

ALIGNMENTS

RESULT 1

T18279

multidrug resistance transport protein - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18279

R:Shaulsky, G.; Loomis, W.F.
submitted to the EMBL Data Library, June 1996

A:Reference number: Z18855

A:Accession: T18279

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1743 <SHA>

A:Cross-references: EMBL:060086; NID:g1399914; PID:g1399915; PIDN:AAB03331.1
C:Genetics:
A:Gene: tagC

Query Match 23.3%; Score 523.5; DB 2; Length 1743;
Best Local Similarity 27.9%; Pred. No. 2.2e-25;
Matches 165; Conservative 81; Mismatches 150; Indels 195; Gaps 24;

QY	19	LYGQGOIVAVADTGLDTR	---NDS-----SMHEAFRGKITALYALGRINNDANGH	68
Db	314	LRGKGQLSIADTGLDGS	HCFFSDSKYPIPLNSVNLNR-KVVTYITTSDDSKVDGH	372
QY	69	GTHVAGSVLG	-----NGSTNGMAPQANLVFQSIMDSGGGLGL--PSNLQTLFQAY	119
Db	373	GPHICGSAAGTPE	DSVNTSFSGLATDAKIAF---FDLASGSSSLTPPSDLKQLOPLY	429
QY	120	SAGARIHTNSWGA	---AVNGAYTTDSRNVDYVRKN-DMTILFAAGNEGPNGTIS--A	172
Db	430	DAGARVHCDWSG	SVSEGYTGSYSSDTASIDDFLTHPDPIILRAGN---NEQYLSLT	486
QY	173	PGTAKNAITVGATENLR	-----PSFGSYADNI-----	199
Db	487	QSTAKNVTVGARHQT	IHENYLTGPNYINVQSSVDINQELICDFDSRYCNYTTAQCCLES	546
QY	200	-----	-----NHVAQFSRSGPTKDGRIKPDVWAPGTFIL	228
Db	547	NATYGLASCCPTLL	RKKSVIDAANTOPLYNENNICSFSKGPHTDGRMKPALVAPGEYIT	606
QY	229	SARSLSA	-----PDSSFWANHDSKYAYMGSTMATPIVAGNVAQLREH-----F	272
Db	607	SARSNGANTDCC	GGSL-PNTNALLA-ISGISMATSFAAATTLILRQVLDVGYPTGSI	664
QY	273	VKNRGITPKP	LLKALIAIAGA-----ADIGLYPNGN-----QMGGRVT	311
Db	665	VESNKLQPTG	SLKALMINNAQLLNGTFLQITSSITYPSNVQNFACASLVQHGAI	724
QY	312	LKSLNVA	YVNESS-----	338
Db	725	MSNWLHV	VNNNNNNKTSBGITKFDGIGGLDLRLVKPNOWKEESLSTGQNTSYCFTYK	784

```

339 -----ATACKPLK---ISLWSDAPASTTASVTILVNDLTLVI-----TAPNGT 378
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
845 PSSSSSSNNIPRVVATLWTDPPSYAGAKFNVLNLDLTMIIYRDNGSTIFYSNQGGS 844
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
379 QYVGNDFTSYNDWGRNENNVFINAPQSGTYTIEQAVNPVGPOTFS 429
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
845 SFLG---LAPTDQT---LNNVEGIVHNPTEPMTYRFWVAGTNVMPGPNFS 889
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 2
T18267
multidrug resistance protein - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18267
R:Shaulsky, G.; Kuspa, A.; Loomis, W.F.
submitted to the EMBL Data Library, January 1995
A:Description: An MDR transporter/serine protease gene is required for prestalk special
A:Reference number: T18850
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1905 <SHA>
A:Cross-references: EMBL:U20432; NID:9664839; PID:9664840; PIDN:AAA62212.1
C:Genetics:
A:Gene: tagB

Query Match 22.1%; Score 497; DB 2; Length 1905;
Best Local Similarity 28.0%; Pred. No. 1.3e-23;
Matches 162; Conservative 72; Mismatches 162; Indels 182; Gaps 22;

QY 19 LYGGQIVAVADTGLDTGR---NDS-----SMHEAFRGKITALYALGRTNNDNTNGH 68
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 376 LRGGQILSLIADTGLDGHCFDSKYPPIFNQVNNHKKVVT---YIYHDNEDYVNGH 432
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 69 GTHVAGSVLNG-----STNKMAPQANLQFQIMDSGGGLGGLPSNLQTLFSAQYSA 121
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 433 GTHVCGSAAAGTPEDSSWAISSFSGLATDAKIAFYD-LSSGSGSEPTPPEDYSQMYKPLLYDA 491
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 122 GARTHNSWGA---AVNGAYTDSNRVDDYVRK-NDMTILFAAGNEGNGGTISAPGTA 176
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 492 GARVHGDSWGSVLOGYGGYSDAGGIDAFLEYEPFSLRAAGN-NELFASLLAQATA 550
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 177 KNATITVGTATENLRPSFGS-----YADNI-----PNCNOGQGRVTLDKSLNVA 199
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 551 KNATITVGAETQAHNVVSDALEYDFSDANFQPCLEFDKKYCNYYTAKCCSEVSNVKG 610
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 200 -----NHVAQFSRGPTKGRIPKPDVWAPGTFILSARSS- 233
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 611 QLCCPASIKQNASDSFTTQPOFYNNENMGFSFKGPTHGRLKPDIVAPGEYITSARSNG 670
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 234 -----LAPDSSEFWANHDSKYAMGTSMATPIVAGNVAOLREHF-----VKNRGI 278
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 671 ENSTDQCGDGL--FNANGLSISGTSMATPLATAATTILRQLVLDGYFTGESVENKL 728
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 279 TPKPSLLKAAALFAGADIGLY-----PNCNOGQGRVTLDKSLNVA 319
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 729 LPTGSLKALMINNAQLNGTVFWSASSTNPSNAIFEQINCANLIOGVALRMN---NWL 785
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 320 YNESS-----SLTSQKAT-----YST-----ATAGK 343
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 786 YKSSNPTPPSRWIGGLGKGNOKATFEWKEDSLSSGLNKSYCFYKPSSSSSSGSGGGGT 845
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 344 P-LKLSLWSDAPASTTASVTILVNDL-----VITAPN--GTQYVGNDFTSYND 391
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 846 PRIVATLWTDPPSYGAKFNVLNLDLNLNDDSDSIITIGSGSLQAGKVAQP--- 902
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 392 NWDGRNNVNFVFINAPQSGTYTIEQAVNPVGPOTFS 429
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 903 ---DTLNNVEGIINPTKAMNYKFTIAGTNVPIGPQKFS 938
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 3
B83891
intracellular alkaline serine proteinase aprX [imported] - Bacillus halodurans (str
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: B83891
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.;
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: B83891
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-444 <STO>
A:Cross-references: GB:AP001513; GB:BA000004; NID:g10174345; PIDN:BA05649.1; GSPDB
A:Experimental source: strain C-125
C:Genetics:
A:Gene: aprX

Query Match 15.5%; Score 347.5; DB 2; Length 444;
Best Local Similarity 29.9%; Pred. No. 6.4e-15;
Matches 107; Conservative 56; Mismatches 110; Indels 85; Gaps 16;

QY 11 DVAOSSYGLYGGQIVAVADTGLDTCRNDSSMHEAFRGKITALY-ALGRTNNDNTNGHG 69
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 135 EVIRNGETLTGKDVTTIAVIDTGI-----YPHEDEGRKAKFVDFVNOREEPYDDNGHG 187
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 70 THVAGSVLNGSTN---KGMAPQANLQFQIMDSGGGLGGLPSNLQTLFSAQYASAGARI 125
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 188 THCAGDAAGNGASSDQYRGPAPEANVIGVKVLNK-QMGSLSEIMQGV----- 235
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 126 HTNSGAAVNGAYTTDSRVN-----DDYVR-----KNDMTILFAAGN 162
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 236 ---EMCIQYNEEHPDPDPIHISMSLGGQALPYNEQEDPMVRIVEEAWNAGTVCVAAAGN 292
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 163 EGPNGGTTISAPCTAKNAITVGATENLRPSFGSVADNHNHVAQFSRGPTKGRIPKPDVMA 222
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 293 SGPDQAOTIASPGVSEKVIIVGALDD-RDITDREDD---VAPFSRGPTIYKPKPDILA 348
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 223 PGTFILSARSSLIAPDSF-----WANHDSKYAMGTSMATPIVAGNVAOLREHFVKNRG 277
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 349 PCNVNIVSLRS---PNSFYDKIQKGRSVGSHYTMMSGTSMATPGCVAGVALMLQH---EPN 402
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 278 ITPKPSLLKAAALFAGADIGLYPNCNOGQGRVTLDKSLNVA---AYNESSLSLSQK 332
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 403 LTPDE--VKTRLM-----EGDRLWA---DRDPNVYAGYISAEAI PNSEE 443
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
A69587
intracellular alkaline serine proteinase aprX - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C:Accession: A69587
R:Kunzt, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.;
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrar
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.;
Iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lard
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Y. M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Port
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Sc
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.;
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uch
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yosh
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus sub
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: A69587
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

```


C; keywords: extracellular protein; hydrolase; serine proteinase
F; 1-32/Domain: signal sequence #status predicted <SIG>
F; 168-423/Domain: subtilisin homology <SBT>

Query Match	12.6%	Score 283;	DB 2;	Length 580;
Best Local Similarity	26.2%;	Pred. No. 1.2e-10;		
Matches 127;	Conservative 55;	Mismatches 147;	Indels 156;	Gaps 24;

Qy	21	GGQITVAVADTGL	---DTGRNDSSMHEAFRGKITALYALGRTNND---	64
Db	168	GGTGVAVIDTGITSHADLNANILAGYDFISDATARDNGRDSNAADEGDWYAANECCA	227	
Qy	65	-----TNGCHGTHVAGSVLNGSTNKGMAFQA-----	99	NLVFQSIMD

```

228 GIPAASSWHGTHVAGTVAATVNTTTCVAGTAYGAKVVPVRVLGRGCGGSLDIADAIVWA 287
Qy 100 SGGGLGGLPSNLQ--TLFSQAYSAGARIHNTSWGAAYNCAYTTDSRNVDDYVVRKNDMTIL 157
Db 288 SGGTVSGTIPANANPAEVIINMSLGGGSCSTTMQN-AINGAVSRGT-----TVV 334
Qy 158 FAAGNEGPNGGTISAPCTAKNATVGTATEN--LRPSFGSYADNINHVAOFSSRGPTKDGR 215
Db 335 VAAGNDASNVSG-SLPANCANVATAATTSAKAKASYNFGTGI----- 377
Qy 216 IKPDVMAPGTFILSARSS--LAPDSEFWANHDSKYAYMGGTSMATPIVAGNVVAQLREHFV 273
Db 378 ---DVSAPGSSILSTLNSGTTTSGS-----ASYASYNGTSMASPHVAGVVALVQS--V 425
Qy 274 KNRGTTPK--PSLLK--AALIAGAADIGLGPNGNGQGRVTLDKSLNVAYNVNESS----- 325
Db 426 APTALTTPAAAVETLLKNTARALPGAC-----SGCGGAGIVNADAATVTA--INGSGGGG 477
Qy 326 -----SLSTSOKATYFTATAGKPLKISLVWSDAPASTTASVTL---VND 367
Db 478 GGNLTNGTPTVTLGNAATCAELNYTITVPAG-----SGILTVTTSGGSGD 523
Qy 368 LDLVI---TAPNGTQYVGNDFTSPTSDNMWGNRVNENFVINAPOSGTYTIEVQAVNVPVG 424
Db 524 ADLYVRAGSAPTDASVT---CRPYRS-----GNAETCTITAP-SGTYTVRLKAYS----- 569
Qy 425 POTFS 429
Db 570 --TFS 572

RESULT 8
S25835
subtilisin (EC 3.4.21.62) precursor - Bacillus sp. (strain TA41)
C:Species: Bacillus sp.
C:Date: 22-Nov-1993 #sequence_revision 20-Feb-1995 #text_change 18-Jun-1999
C:Accession: S25835
R:Davail, S.; Feller, G.; Narinx, E.; Gerday, C.
Gene 119, 143-144, 1992

```

Db 135 GAGINIAVLDTGNTNHPDLSNNVEQCKD-----FTVGTINFTDNCSTDRQGHGTHVAGSA 189
QY 77 LCGNSTNK---GMAPOANL-VFQSIMDSGGGLG-GLPSNLTQTLFSQAYSAGARLHTN-SW 130
Db 190 LANGGTGSGVYGAPADLWAKVIGDDSGGYADDAEAIIRHAGDQATALNTKVITNMSL 249
QY 131 GAAVNGAYTTDSRNVDYVRKNDMTILFAAGNECPNGGTISAPGTAKNATITVGATENLRP 190
Db 250 GSSGESSLLIT---NAVDAAYDKGVLIITAAAGNSGPKPGSGYGCALVNVAVALENTIQ 306
QY 191 SFGSYADNINHVAQSSRSGPTKDG-----RKPDYMACTFTILSARSSSLAPDSSFWANH 244
Db 307 N-GTY-----RVADFSSRHKRTAGDYVIOKGDVEISAPCAAYST-----W-F 348
QY 245 DSKYAYMGTSMATPIVAGNVAQL 268
Db 349 DGGYATISGTSWASPHAGLAAKI 372

RESULT 9
G83753
subtilisin-type proteinase (EC 3.4.21.-) vpr precursor [similarity] - Bacillus halodurans
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: G83753
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: AB36650; MUID:20512582; PMID:11058132
A:Accession: G83753
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-799 <STO>
A:Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BAB04550.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: vpr
C:Superfamily: microbial serine proteinase vpr; subtilisin homology
C:Keywords: hydrolase; serine proteinase
F:1-29/Domain: signal sequence #status predicted <SIG>

Query Match 12.5%; Score 280; DB 2; Length 799;
Best Local Similarity 24.7%; Pred. No. 3e-10;
Matches 117; Conservative 50; Mismatches 135; Indels 172; Gaps 16;

QY 18 GLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITAYALGRNNDT-----NGH 68
Db 171 GYTCEGITVAIDRGVDYTHPD--LVHAF-GDYKQWDFIDNDDPQETPPGDPGRGIETH 227
QY 69 GTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGSLPSNLTQTLFSQAYSAGARIHTN 128
Db 228 GTHVAGTVAANGLI-KGVAPDANLLAYRVLGPGG--RGSTAGVIAGIERAVQDGADIMNL 284
QY 129 SWGAAGVNGAYTTDSRNVDYVRKNDMTILFAAGNECPNGGTISAPGTAKNATITVGATENL 188
Db 285 SLGNTLNDPDPFATISAL-DWMAEGCVAVTNGNSGPNNTVGSPTSRDALSVCAT--- 340
QY 189 RPSPGSY-----ADNINH----- 201
Db 341 RLPYNYKASVFTSDGIDYPSADIMGPPSDEELLELDGETYEYAFAGLGKPCDFGVDVE 400
QY 202 ----- 201
Db 401 GKIALIVRGELPFVEKAENAKAAGAVGAIYNNVAGVQTPVGLAIPITMLSNEDCLKMR 460
QY 202 -----VAQFSSRGPT-KDGRIPKPDVMAAGPFFILSARSSSLAPD 237
Db 461 NELENGONTVTFSTFEDKLVGETVADFSSRGPPVMTWMIKPDVSPAGVAIVSTIPTHOPD 520
QY 238 SSTWANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAAALIAAGADI- 296
Db 521 DPY-----GYGSRQGTSMASPHVAGAAALLLEAH-FNWGV----DHVKAALMNTAENLV 569

QY 297 ---GLGYPNGNOGWRVTLDKSLNVAIVNNESSLSTSQKATY-SFTATAGKPLK 346
Db 570 DENGNRYPHNTQAG-----SIRIVDAIESETLVTPGSHSFGTFTTKERKQOVE 617

RESULT 10
T29050
surface layer-associated STABLE proteinase - Staphylothermus marinus
N:Alternate names: hyperthermostable proteinase
C:Species: staphylothermus marinus
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T29050
R:Mayr, J.; Lupas, A.; Kellermann, J.; Eckerskorn, C.; Baumeister, W.; Peters, J.
Curr. Biol. 6, 739-749, 1996
A:Title: A hyperthermostable protease of the subtilisin family bound to the surface
A:Reference number: Z20559; MUID:96385442; PMID:8793300
A:Accession: T29050
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-1345 <NAY>
A:Cross-references: EMBL:U57968; NID:g1374755; PID:g1374756; PIDN:AAB02323.1
A:Experimental source: strain F1
C:Function:
A:Note: stoichiometric S-layer component

Query Match 12.4%; Score 279.5; DB 2; Length 1345;
Best Local Similarity 29.9%; Pred. No. 6.4e-10;
Matches 99; Conservative 45; Mismatches 106; Indels 81; Gaps 14;

QY 46 FRGKITAYALGRTNNDTNGHGHVA-----GSVL-----GNSGTNK--GMAPOANLV 93
Db 445 YQGRYAL-----VSDFHGHTSVATVIASRGVLYDLYGDGKLYRIMGVAPGAKI- 495
QY 94 FQSIMDSGGGLGSLPSNLTQTLFSQAYSAG-----ARHNTSW 130
Db 496 -----AGGDALWLLGNILVL--EAWLAGFNIVTEEDGYVYVLSLDPPGPHRADILSNW 546
QY 131 GAAVNGAYTTDSRNVD-----DVKRNDMTILFAAGNECPNGGTISAPGTAK 177
Db 547 GSTYINFWLQOFGIDYRSSFMDEILAIRNTLIGDHVTIVFAAGNEGPGYSSNGAPGTGL 606
QY 178 NAITVGATE--NLRPSFG---SYADNINHVAQSSRSGPTKDGRIPKPDVMAAGPFTILSARS 232
Db 607 LVITAGASTLWDYTRIYGPEGYAD---EVIPFSSRGPTGGQYPPKPDIVNIGAFEWASTR 663
QY 233 SLAPDSSFANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAAALIAAG 292
Db 664 TI-DGRGYGAQPD-----VFGTSEATPVTSGTLALVFOAYKEVYNTTDPDPTAKILKSS 718
QY 293 AADIGLGYPNNGOGWRVTLDKSLNVAIVNE 323
Db 719 AKDI--WYPAFSGSGRVDALKAADTVFISE 747

RESULT 11
S23407
subtilisin (EC 3.4.21.62) l precursor - Bacillus sp. (strain TA39)
C:Species: Bacillus sp.
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 18-Jun-1999
C:Accession: S23407
R:Narinx, E.; Davail, S.; Feller, G.; Gerday, C.
Biochim. Biophys. Acta 1131, 111-113, 1992
A:Title: Nucleotide and derived amino acid sequence of the subtilisin from the antar
A:Reference number: S23407; MUID:92256481; PMID:1581352
A:Accession: S23407
A:Molecule type: DNA
A:Residues: 1-420 <NAR>
A:Cross-references: EMBL:X62369; NID:g40200; PIDN:CAA44227.1; PID:g40201
C:Genetics:
A:Gene: subtil
C:Superfamily: subtilisin; subtilisin homology
C:Keywords: extracellular protein; hydrolase; serine proteinase

F:1-23/Domain: signal sequence #status predicted <SIG>
 F:24-111/Domain: propeptide #status predicted <PRO>
 F:112-420/Product: microbial serine proteinase #status predicted <MAT>
 F:136-374/Domain: subtilisin homology <SBT>
 F:145,185,360/Active site: Asp, His, Ser #status predicted

Query Match 12.3%; Score 277; DB 1; Length 420;
 Best Local Similarity 31.2%; Pred. No. 1.9e-10;
 Matches 98; Conservative 44; Mismatches 120; Indels 52; Gaps 16;

QY 21 GQGIIVADVADGLDTRNDSSMHEAFRGKITAL--YALGRT---NNAANDNGHGHVAGS 75
 DB 136 GGGINIAVLDGVTN-----HPLDRNVEOCKDFTVGTITVNNSTDRQGHGHVAGS 189
 QY 76 VLNGSTNKK---GMAPQANL-VFQSIMDSGGGLG-GLPSNLTQFSQAYSAGARIHTN-S 129
 DB 190 ALADGGTGGNGYGVAPDADLWAYKVLGDDGGYADDAIAAIRHAGDOATALNTKVVINMS 249
 QY 130 WGAANGAYTTDSRNVDDYVRKNDMTILFAACNEGPNGGTISAPGTAKNAITVGATENLR 189
 DB 250 LGSSGESLITNAVN---YSYNGKGLIITAAAGNSGPGYQSGIPGALVNAVVAALEN-K 305
 QY 190 PSFGSYADNINHVAFSSRGPT-KDG-----RIKPDVMAPGTFILSARSLAPDSFEW 243
 DB 306 VENGTY-----RVADFSGYSWTDCDYAIQKGDVEISAPGAAYST-----W-- 348
 QY 244 HDSKYAMGGTSMATPIVAGNVAQLREHFVKRGLTPKPSLLKAAALIAGAADIGLGPNG 303
 DB 349 FDGGTATTSYGSPHAAAGLAANKIAQWPSASNVDRVGLQYRAY---ENDILSGYYAG 405
 QY 304 -----NOGWGRVTL 312
 DB 406 YGDDEFAFGFGFATV 419

RESULT 12

JC4908
 alkaline serine proteinase (EC 3.4.-.-) I precursor - Alteromonas sp.
 A:Alternate names: subtilase
 C:Species: Alteromonas sp
 C:Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 08-Oct-1999
 C:Accession: JC4908
 R:Tsujibo, H.; Miyamoto, K.; Tanaka, K.; Kaidzu, Y.; Imada, C.; Okami, Y.; Inamori, Y.
 Biosci. Biotechnol. Biochem. 60, 1284-1288, 1996
 A:Title: Cloning and sequence analysis of a protease-encoding gene from the marine bacterium *Alteromonas*
 A:Reference number: JC4908; MUID:97141200; PMID:8987544
 A:Accession: JC4908
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-715 <TSU>
 A:Cross-references: DBJ:D38600; NID:g1536787; PIDN:BAA18912.1; PID:d1019647; PID:g21602
 A:Experimental source: strain O-7
 C:Comment: This enzyme belongs to class I subtilisin-like family. It is a chelator-sensit

C:Genetics:
 A:Gene: aprI
 C:Superfamily: subtilisin homology
 C:Keywords: hydrolase
 F:1-40/Domain: signal sequence #status predicted <SIG>
 F:41-150/Domain: amino-terminal propeptide #status predicted <ATP>
 F:151-496/Product: alkaline serine protease I #status predicted <MAT>
 F:182-452/Domain: subtilisin homology <SBT>
 F:497-715/Domain: carboxyl-terminal propeptide #status predicted <CTP>
 F:239-294,335-372,478-481/Disulfide bonds: #status predicted

Query Match 11.9%; Score 267; DB 2; Length 715;
 Best Local Similarity 25.9%; Pred. No. 1.7e-09;
 Matches 124; Conservative 51; Mismatches 188; Indels 136; Gaps 23;

QY 21 GQGIIVADVADTG----LDTGRNDSMHEAFRGKITALYALGRTNNAND-----TNG----- 67
 DB 182 GGGVVAVLDGTGRPHLDLADNLPGYDMISNTFVNDGGARDNDARPDGAVTRGCGT 241
 QY 68 -----HGHVAG---SVLNGSTNKGMAQANLVFQSIMDSGGGLGLP 108

DB 242 DSSGQPVPRADODSSWHGTHVAGTAAVTNNGEGVAGVADAKVVPVRVL---GKCGLT 298
 QY 109 SNLTQFSQAYSAGARIHTNSWGAIV-----NGAYTTDSRNVDDYVRKNDMTILFPAAG 161
 DB 299 SDIADGLIWAAGSDRVFANANPAVVINNSLGGGACSATTOINAINOARNGTVIVITAA 358
 QY 162 NEGPNGGTISAPGTAKNAITVGAT--ENLRPSFGSYADNINHVAFSSRGPTKDRIKPD 219
 DB 359 NDNDNSANY-PCNGGVNVASVGRDGRSAYISNYGANI-----D 398
 QY 220 VMAPGTFILSARSLAPDSFEWHDSD-----KYAYMGTSMATPIVAGNVAQLR-- 269
 DB 399 VAAPG---GAQSFADDPGEGILLSTHNSGCGAPSNDSYHYSGTSMAPHAAPHVAGVAALIKQA 454
 QY 270 -----EHFVKN--RGITPKPSLLKAAALIAGAADI--GLG---YPCNGMQGWRVTL 312
 DB 455 KPSATPDEVETTLKNTTRSFAGSCNCGTGVDDAAANALGDDVVTPTGN-----TL 508
 QY 313 DKSLNVAVYNSSSLTSQKATYSFTATAGKPLKISLVWSDAPASTTASVTL---VNDLD 369
 DB 509 ED--GVAKTGLSGAAGSNQ--PFTFDVPAGK-----TNVFTMSGGTGDAD 550
 QY 370 LVITAPNGTQYVGNDEFTSPYNDWDGR-----NNVENVFINAPOSQTYTIEVQAVNPVG 424
 DB 551 LVYVKL--GSQ-----PTSSSYDCRPYEGGNAEVCDFDAPQAGTVHVMINGYKAYSG 599

RESULT 13

A72647
 probable surface layer-associated STABLE proteinase APE0607 - Aeropyrum pernix (strain ATCC 49239)
 C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
 C:Accession: A72647
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-do, K.; Tanaka, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, H.; Yamazaki, S.
 DNA Res. 6, 83-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, *Aeropyrum pernix* strain A72647
 A:Reference number: A72647; MUID:99310339; PMID:10382966
 A:Accession: A72647
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1331 <KAW>
 A:Cross-references: DBJ:AP000060; NID:g5104188; PIDN:BAA79577.1; PID:d1043363; PID:g21602
 A:Experimental source: strain K1
 C:Genetics:
 A:Gene: APE0607

Query Match 11.7%; Score 263; DB 2; Length 1331;
 Best Local Similarity 23.6%; Pred. No. 7.2e-09;
 Matches 111; Conservative 65; Mismatches 158; Indels 136; Gaps 18;

QY 8 KVADVAQSSYGLXGQGIIVADVADTGLDTRNDSSMHE-AFRGKITALYA----- 55
 DB 346 VYADLS-TAYLFP---LKALSDTGSMISGEPDPSLLDLSEFADETAPASYSEVLARDFTGD 400
 QY 56 -----LGRTNAN-----DTNGH 68
 DB 401 GVNDFSAGALAGWTYDWVGLLTGSEVNLGRLGFDYAGLVLPGLDPQGRWVSILYDTLAH 460
 QY 69 GTHVAGSVLNGSGTN-----KGMAPQANLVFQSIMDSGGGLGLPLSNLTLSQA 118
 DB 461 GTSVAIVIASRGNVEFFNLGYIETSLRGVAPGAKTA-----AGGSF-----LINVFVAQL 509
 QY 119 YSAG-----ARIHTNSWG---AAVNGAYT---TDSRNVDDY-VRKNDMT 155
 DB 510 FLSGFEPQDSPLNVTYTGHEQVDVINNSWGNYSIALRGFLTGADDTATIEDIVSAGTV 569
 QY 156 ILFAAGNEGNGGTISAPGTAKNAITVGATE--NLRPSFGSYADNINHVAFSSRGPTKD 213
 DB 570 IVHAMGNGGPGYGTATTPGAGSLIISVGASTLFDYRPFYGYLPSGCGDVISWDRGFSQI 629
 QY 214 GRIKPDVMAQPTFILSARSLAPDSFEWHDSDSKYAYMGGTSMATPIVAGNVAQLREHFV 273

```
Db      630 GVAKPDVVMNIGSAWAG-----VPVLTGLNGSLAFDIFGCTSEATPMTSGSVALVISAQ 685
QY      274 KNRGITPKSILKAAALAGAADIGLGPNGQGWGRVTLDKSLNV-----AY 320
Db      686 QAFGAKPSGLVKALKLKSTARTGA--DAFTQGSGQVDYRVAVKAVLEGVPIALSTSVY 743
QY      321 VNESSLSSTQKATYSFTATACKPLKISLWSDA--PASITASVTLVNDL 368
Db      744 ENVYSLLS-----GYSYPLAPNPVEDTQIYPCVLKPGETAVETLVKLTL 798

RESULT 14
C84120
subtilisin-type proteinase (EC 3.4.21.-) BH3763 precursor [similarity] - Bacillus halodu
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: C84120
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: C84120
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-757 <STO>
A:Cross-references: GB:AP001519; GB:BA000004; NID:g10176109; PIDN:BA07482.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH3763
C:Superfamily: microbial serine proteinase vpr; subtilisin homology
C:Keywords: hydrolase; serine proteinase
F:1-20/Domain: signal sequence #status predicted <SIG>

Query Match      11.5%; Score 259.5; DB 2; Length 757;
Best Local Similarity 22.6%; Pred. No. 5.6e-09;
Matches 119; Conservative 58; Mismatches 155; Indels 195; Gaps 20;

QY      8 VKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRG-----KITALYALG 57
Db      119 VRGMLDEEGVHLTGKVKVAVDTGIDYTHPD--LOSSYKGGYDFVDYDDDPMTIASQG 176
QY      58 RTNNANDTNGHGHVAGSVLNGSTNKGMAQANLVFQSIIMDSGGGLGGLPSNLQTLFSQ 117
Db      177 -----PPTLHGTHVSGIIAANGQV-KGVAPEAIIYAYRALPGC--QGTEQVIAAIEK 227
QY      118 AYSAGARIHTNSGWAAYNCAYTTDSRVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAK 177
Db      228 AVEDGVDVNLNLGNITVNGPDWPTSLDAAEVGVAVT-SNGNSGPNMWTYGSPTSK 286
QY      178 NAITVGAT-----EN--LRPSFGS----- 194
Db      287 KAISVGASAPLNTPLYLAFGENEISLIPFSGGLPWAFKRLDPLMIDVGYGTEKEGVD 346
QY      195 -----YADNINH----- 201
Db      347 AEGKVLIKRGVMPTEKVMHAAKARGVLIYNNTPPTGMEGGVNIPIVVSITREDG 406
QY      202 -----VAQFSRGP-TKDGRIKPDVMACTILSRSS 233
Db      407 EPLLQLEQLQKNKELTLRTYRKEDFVALFSRGPVTHWDVKPDVAPGVSI-----DS 462
QY      234 LAPDSFWANHDSKYAYMGSTMATPIVAGNVAOLREHFVKNRGITPK--PSLLKAAALIA 291
Db      463 TIPNNG-----YLGLNGTSMAPHVAGAAALIKQ-----AHPWTPEQVKAALMN 507
QY      292 GAADI-----GLYPNGNQGRVTLDKSLNVAVYVNESSLSSTQKATYSFTATAGPLKI 347
Db      508 TAKKLVDQEGVPHEHTEQAGRIQVDKAV-----AATSLVYPCALSEFGK---- 551
QY      348 SLVWSDAPASTTASVTIL-VNLDLVLITAPNGTOYVGNDFTSYNDNW 393
Db      552 ---WSKDDLUREKRPVTLTIENHDTV----KRTYHISPPFDVDPGV 591
```

RESULT 15

```
A35742
aqualysin (EC 3.4.21.-) I precursor - Thermus aquaticus
C:Species: Thermus aquaticus
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: A35742; S00620; S00324
R:Terada, I.; Kwon, S.T.; Miyata, Y.; Matsuzawa, H.; Ohta, T.
J. Biol. Chem. 265, 6576-6581, 1990
A:Title: Unique precursor structure of an extracellular protease, aqualysin I, with
A:Reference number: A35742; MUID:90216674; PMID:2182621
A:Accession: A35742
A:Molecule type: DNA
A:Residues: 1-513 <TER>
A:Cross-references: GB:J90108; GB:D90108; GB:J05414; NID:g217171; PIDN:BAAL4135.1; I
A:Note: The authors translated the codon CTG for residue 470 as Val, and GGT for res
R:Kwon, S.T.; Terada, I.; Matsuzawa, H.; Ohta, T.
Eur. J. Biochem. 173, 491-497, 1988
A:Title: Nucleotide sequence of the gene for aqualysin I (a thermophilic alkaline s
A:Reference number: S00620; MUID:88225062; PMID:3286255
A:Accession: S00620
A:Molecule type: DNA
A:Residues: 75-442 <KWO>
A:Cross-references: EMBL:X07734; NID:g48069; PIDN:CAA30559.1; PID:g602091
A:Note: part of this sequence, including the amino and carboxyl ends of the mature
R:Matsuzawa, H.; Tokugawa, K.; Hamaaki, M.; Mizoguchi, M.; Taguchi, H.; Terada, I.;
Eur. J. Biochem. 171, 441-447, 1988
A:Title: Purification and characterization of aqualysin I (a thermophilic alkaline
A:Reference number: S00324; MUID:88151937; PMID:3162211
A:Accession: S00324
A:Molecule type: protein
A:Residues: 128-170 <MATS>
C:Superfamily: subtilisin; subtilisin homology
C:Keywords: extracellular protein; hydrolase; serine proteinase
F:1-14/Domain: signal sequence #status predicted <SIG>
F:15-127/Domain: propeptide #status predicted <PRO>
F:128-408/Product: aqualysin I #status experimental <MAT>
F:157-364/Domain: subtilisin homology <SBT>
F:255-257, 281-283/Region: S1 specificity crevice #status predicted
F:409-513/Domain: carboxyl-terminal propeptide #status predicted <CPR>
F:166,197,349/Active site: Asp, His, Ser #status predicted

Query Match      11.5%; Score 257.5; DB 1; Length 513;
Best Local Similarity 26.0%; Pred. No. 4.5e-09;
Matches 117; Conservative 44; Mismatches 144; Indels 145; Gaps 23;

QY      16 SYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALLY-ALGRTNNANDTNGHGHVAG 74
Db      152 TYTATGRGVNVVVDIGTIRT-----THREFGGRARVGYDALG--GNGQDCNGHGHVAG 203
QY      75 SVLGNSTNKGMAQANLVFQSIIMD--SGGLGGLPSNLQTLFESQAYSAGARITN---- 128
Db      204 TI---GGVTYGVAKAVNLYAVRVLDCNGSGTSGVIAGVDW-----TRNHRPVA 252
QY      129 --SWGAAVNGAYTTDSRVNVDYVRKN---DMTILFAAGNEGPNGGTISAPGTAKNALITVG 183
Db      253 NWSLGGGVSTA-----LDNAVKNLSIAGVVYVAAAGNDNANACNYS-PARVAEALTVG 304
QY      184 AT--ENLRPSFGSYADNINHVAFQFSRGPDKGRTPKDVYMAPGTFILSARSSLAPDSSF 241
Db      305 ATTSDDARASTSNYSGCV-----DLFAPGASIPSA-----W 335
QY      242 ANHLSKAYAMGCTSNATPIVAGNVAOLREHFVKNRGITP---KPSLLKAAALIAAGADIGL 298
Db      336 YTSOTATOTLNGTSMATPHVAG-VAAL--YLEQNPSPASPASVASALINGATTGRLSGIGS 392
QY      299 GYPNGNQGRVTLDKSLNVAVYVNESSLSSTQKATYSFTATAGKPLKISLVWSDAPAST 358
Db      393 GSPN-----RLLYSLLSSGS-----GSTAPCTS 415
QY      359 ----TASVTLVNDLVLITAPNGTOY-----VGNDFTSYNDNWGNR---- 397
```

Db 416 CSYYTGSLSGPDYNE---QPNGTYYYSPAGTHRAWLRGPAGTDF-DLYLWRWDGSRWLT 471
QY 398 -----NVENVFINAPQSGTYTIEVQAYN 420
Db 472 VGSSTGPTSEESLSYSGTAGYVYLWRIYAYS 501

Search completed: July 25, 2003, 19:02:03
Job time : 20.0185 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 18:56:30 ; Search time 9.01038 seconds
(without alignments)
2265.120 Million cell updates/sec

Title: US-09-985-689A-1

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSVGLY.....EVQAYNVPGPQTFSLAIVN 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	523.5	23.3	1743	1	TAGC_DICDI
2	497	22.1	1905	1	TAGE_DICDI
3	310.5	13.8	806	1	SUBV_BACSU
4	304.5	13.6	1398	1	PLS_PYRFO
5	283	12.6	580	1	EXPR_XANCP
6	277	12.3	420	1	SUBT_BACSP
7	257.5	11.5	513	1	AQLI_THEAQ
8	253.5	11.3	894	1	WPRA_BACSU
9	246.5	11.0	402	1	ALP_CEPAC
10	246.5	11.0	534	1	PROA_VIBAL
11	242.5	10.8	401	1	THES_BACSP
12	240	10.7	269	1	SUBS_BACLE
13	240	10.7	380	1	ELVA_BACAO
14	240	10.7	380	1	ELVA_BACCS
15	239	10.6	269	1	PRTM_BACSP
16	235.5	10.5	382	1	SUBT_BACAM
17	235	10.5	378	1	ELVA_BACSP
18	232	10.3	269	1	SUBB_BACLE
19	231	10.3	321	1	ISP_BACCS
20	230.5	10.3	379	1	SUBT_BACLI
21	230	10.2	404	1	SMP1_MAGPO
22	230	10.2	1181	1	SCA2_STRPY
23	227	10.1	1167	1	SCA1_STRPY
24	225.5	10.0	381	1	SUBN_BACNA
25	225.5	10.0	1052	1	MSIP_CRIGR
26	225.5	10.0	1052	1	MSIP_HUMAN
27	225.5	10.0	1052	1	MSIP_MOUSE
28	225.5	10.0	1052	1	MSIP_RAT
29	223.5	9.9	381	1	SUBT_BACSA
30	223.5	9.9	381	1	SUBT_BACST
31	223.5	9.9	381	1	SUBT_BACSU
32	221.5	9.9	275	1	SUBT_BACPU
33	221.5	9.9	530	1	HLX_HALL17

ALIGNMENTS

RESULT 1

ID	TAGC_DICDI	STANDARD	PRT	1743 AA
AC	Q23868			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Prestalk-specific protein tagC precursor (EC 3.4.21.-)			
GN	TAGC			
OS	Dictyostelium discoideum (Slime mold).			
OC	Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.			
OX	NCBI_TaxID=44689;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=AX4;			
RX	MEDLINE=97140317; PubMed=8986798;			
RA	Shaulsky G., Escalante R., Loomis W.F.;			
RT	"developmental signal transduction pathways uncovered by genetic suppressors.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 93:15260-15265(1996).			
CC	FUNCTION: INTERCELLULAR COMMUNICATION VIA TAGC MAY MEDIATE INTEGRATION OF CELLULAR DIFFERENTIATION WITH MORPHOGENESIS (BY SIMILARITY).			
CC	FUNCTION: IN THE N-TERMINAL SECTION; BELONGS TO PEPTIDASE FAMILY S8.			
CC	SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS). MDR SUBFAMILY.			
CC	SIMILARITY: STRONG, TO TAGC.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL: U60086; AAA03331.1; -			
DR	PIR: T18279; T18279.			
DR	DictyDB: DD02060; tagC.			
DR	InterPro: IPR003593; AAA_Atptase.			
DR	InterPro: IPR001140; ABC_TM_transpt.			
DR	InterPro: IPR0003439; ABC_transporter.			
DR	InterPro: IPR000209; Peptidase_S8.			
DR	Pfam: PF00664; ABC_membrane; 1.			
DR	Pfam: PF00005; ABC_tran; 1.			
DR	Pfam: PF00082; Peptidase_S8; 1.			
DR	PRINTS: PF00723; SUBTILISIN.			
DR	SMART: SM00382; AAA; 1.			
DR	PROSITE: PS00211; ABC_TRANSPORTER_1; 1.			
DR	PROSITE: PS00893; ABC_TRANSPORTER_2; 1.			
DR	PROSITE: PS00136; SUBTILASE_ASP; FALSE_NEG.			
DR	PROSITE: PS00137; SUBTILASE_HIS; 1.			
DR	PROSITE: PS00138; SUBTILASE_SER; FALSE_NEG.			
KW	Hydrolase; Serine protease; ATP-binding; Transport; Transmembrane; Signal.			

P00781 bacillus li
P16396 bacillus su
P58502 pyrococcus
P41363 bacillus ha
P04072 thermoactin
P20015 tritirachiu
P29139 paenibacill
Q03420 trichoderma
P11018 bacillus su
P16397 bacillus su
P29138 metarhizium
P23653 tritirachiu

34 216.5 9.6 274 1 SUBD_BACLI
35 216.5 9.6 645 1 SUBE_BACSU
36 215 9.6 422 1 TKSU_PYRKO
37 214.5 9.5 361 1 ELYA_BACHD
38 214 9.5 279 1 THET_THEVU
39 214 9.5 293 1 PRTH_TRIAL
40 214 9.5 326 1 ISP_PAEPO
41 214 9.5 409 1 ALP_TRIHA
42 211 9.4 319 1 ISPI_BACSU
43 209.5 9.3 1433 1 SUBF_BACSU
44 208.5 9.3 388 1 CUDP_METAN
45 208 9.3 387 1 PRTR_TRIAL

"The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.";
 RL Nature 390:249-256(1997).
 CC -!- FUNCTION: NOT REQUIRED FOR GROWTH OR SPORULATION.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- PTM: PROBABLY UNDERGOES C-TERMINAL PROCESSING OR PROTEOLYSIS.
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M76590; AAA22881.1; -;
 CC EMBL; X73124; CAA51601.1; -;
 CC EMBL; Z99123; CAB15835.1; -;
 CC PIR; A41341; A41341.
 CC HSP; P00782; 2S8T.
 CC SUBTIL; BG10591; vpr.
 CC InterPro; IPR003137; PA.
 CC InterPro; IPR000209; Peptidase_S8.
 CC Pfam; PF02225; PA; 1.
 CC Pfam; PF00082; Peptidase_S8; 1.
 CC PRINTS; PR00723; SUBTILISIN.
 CC PROSITE; PS00136; SUBTILASE_ASP; 1.
 CC PROSITE; PS00137; SUBTILASE_HIS; 1.
 CC PROSITE; PS00138; SUBTILASE_SER; 1.
 CC Hydrolyase; Serine protease; zymogen; Signal; Complete proteome.
 KW SIGNAL 1 28
 FT SIGNAL 1 28
 FT PROPEP 29 160
 FT CHAIN 161 806
 FT ACT_SITE 189 189
 FT ACT_SITE 233 233
 FT ACT_SITE 534 534
 FT ACT_SITE 806 AA; 85608 MW; F984E3BF0B869DDD CRC64;
 SQ SEQUENCE 806 AA; 85608 MW; F984E3BF0B869DDD CRC64;
 Query Match 13.8%; Score 310.5; DB 1; Length 806;
 Best Local Similarity 23.5%; Pred. NO. 6.6e-13;
 Matches 129; Conservative 58; Mismatches 157; Indels 205; Gaps 18;
 QY 18 GLYGGQIVAVADTGLDTGR-----NDSSMHEAFRGKITALYALGRTN 60
 DB 177 GTYGGKGLKVAIDRGVEYNHDPDLKKNFGQYGYGVDFVNDYDKPTG-----D 225
 QY 61 NANTDNGHGHVAGSVLGNSTNGKMPAQANLVFQSIMDSGGGLGLPLSNLQTLFSQAYS 120
 DB 226 PRGEATDGHVAGTVAANG-TIKGVAPDATLLAYRVLPGG--SGTTENVIAGVERAVQ 282
 QY 121 AGARIHNSGAANGAYTTDSRNVDYVRKNDWTLILFAAGNEGNGGTISAPGTAKNAI 180
 DB 283 DGADVMNLSLGNLNPDPWATSTAL-DWAMSEGVAVTNGNSGPNGTGVSPTGREAI 341
 QY 181 TVGATE---NLRFSEGSY-----195
 DB 342 SVGATQLPLNEYAVTFGSYSYSAKVMYKEDDVKALNNKEVELVEAGIGKDFEGKDLT 401
 QY 196 -----ADNI-----199
 DB 402 GKVAVKRGSTAFVDBKADNARKAGAIMGVVYNNLSGEIANVPMSVPTIKLSLEDGEKL 461
 QY 200 -----NHVAQSSRGPTKD-GRIKPDVMAPGFFILSARSLAP 236
 DB 462 VSALKAGETKTTFKLTIVSKALGEQVADFSSRGVPMDWIKPDISAPGVNIVSTIPTHDP 521
 QY 237 DSSEWANHDSKYAYMGTSMAPIVAGNVAGNVAQLREHFVKNRGTTPKPSL--LKAALTAGAA 294
 DB 522 D-----HPYGYGSKQGTSMASPHIAGAVAIKQ-----AKPKWSVEQIKAAIMNTAV 568
 QY 295 DI-----GLGYPNGNOGWGRVTLDKSLINVAIVYNESSLSQKATYSFTAGPKLSLV 350

Db 569 TLKDSGEYYPHNAQAGSARI-----MNAIKADSLVSPGSYSY-----GTFLK----- 612
 QY 351 WSDAPASTTASVTLVNDLVLVITAPNGTQVYVGNDETSPYNDWDG--RNNVENVFNAQ 408
 Db 613 -ENGTNETFTIENG-----SSIRKSYTLEYSFNGSGISTSGTSRVVIPAHO 660
 QY 409 SGTYTIEVQ 417
 Db 661 TGRKATAVK 669
 RESULT 4
 PLS_PIRFU STANDARD: PRT: 1398 AA.
 ID PLS_PIRFU AC P72186;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Pyrolysin precursor (EC 3.4.21.-).
 GN PLS OR PF0287.
 OS *Pyrococcus furiosus*.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC *Pyrococcus*.
 OC NCBI_TaxID=2261;
 RN [1]
 RC SEQUENCE FROM N.A., SEQUENCE OF 150-184, AND CHARACTERIZATION.
 RP STRAIN=Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
 RX MEDLINE=96355370; PubMed=8702780;
 RA Voorhorst W.G.B., Eggen R.I.L., Geerling A.C.M., Platteuw C.,
 RA Siezen R.J., de Vos W.M.;
 RA "Isolation and characterization of the hyperthermostable serine
 RT protease, pyrolysin, and its gene from the hyperthermophilic archaeon
 RT *Pyrococcus furiosus*.";
 RL J. Biol. Chem. 271:20426-20431(1996).
 RN [2]
 RC SEQUENCE FROM N.A.
 RP STRAIN=Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
 RX Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
 RA "The complete sequence of the *Pyrococcus furiosus* genome.";
 RT Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RC CHARACTERIZATION, AND 3D-STRUCTURE MODELING.
 RP MEDLINE=21079021; PubMed=11210516;
 RX de Vos W.M., Voorhorst W.G.B., Dijkgraaf M., Kluskens L.D.,
 RA van der Oost J., Siezen R.J.;
 RA "Purification, characterization, and molecular modeling of pyrolysin
 RT and other extracellular thermostable serine proteases from
 RT hyperthermophilic microorganisms.";
 RL Meth. Enzymol. 330:383-393(2001).
 CC -!- FUNCTION: Has endopeptidase activity toward caseins, casein
 CC fragments including alpha-S1-casein and synthetic peptides.
 CC -!- SUBCELLULAR LOCATION: Cell-envelope associated.
 CC -!- PTM: LWM pyrolysin seems to be produced by autoprolytic
 CC activation of HWV pyrolysin.
 CC -!- PM: Glycosylated.
 CC -!- MISCELLANEOUS: Thermostable; high activity at 95 degrees Celsius.
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U55835; AAB09761.1; -;
 CC EMBL; AE010153; AAL80411.1; -;
 CC PIR; T28159; T28159.
 CC HSP; Q45670; 1DB1.
 CC MEROPS; S08.100; -;
 CC InterPro; IPR000209; Peptidase_S8.

InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
KW Hydroxylase; Serine protease; Zymogen; Signal; Complete proteome.
FT SIGNAL 1 32
FT PROPEP 33 2136
FT CHAIN 2137 580
FT ACT_SITE 177 237
FT ACT_SITE 237 237
FT ACT_SITE 409 409
FT DISULFID 225 273
FT DISULFID 315 352
FT DISULFID 450 454
SQ SEQUENCE 580 AA; 57228 MW; 8C9A2CAB4E7F47CB CRC64;

Query Match 12.6%; Score 283; DB 1; Length 580;
Best Local Similarity 26.2%; Pred. No. 2.7e-11;
Matches 127; Conservative 55; Mismatches 147; Indels 156; Gaps 24;

QY 21 GGGQIVAVADTGL-----DTGRNDSSMHEAFRGKITALYALGRTNND-----64
Db 168 GSGTVAVVDTGITSHADLNAILAGYDFISDATTARDGNGRDSNADEGDWYAANECCA 227

QY 65 -----TNGHGTTHVAGSVLNGSTNKGMAPOA-----NLVFQSIMD 99
Db 228 GIPAASSSWHGTTHVAGVAAVTNNTTGVAGTAYGAKVVPVRLGKCGSLSDIADAIWVA 287

QY 100 SGGGLGLPSNLQ--TLFSQAYSAGARHTHTNSGAAVNGAYTTDSRNVDVYVRKNDMTIL 157
Db 288 SGGTSGIPANANPAEVIINSLGGGCGSSTTMQN-AINGAVSRGT-----TVV 334

QY 158 FAAGNEGPNGGTSAPGTAKNAITVGATEN--LRPFSQSYADNINHVAFQSSRGPTKQGR 215
Db 335 VAAGNDASNSVG--SLPANCANVIAVAATTSAGAKASYNFGTGI-----377

QY 216 IKPDVMAPGTFILSARSS--LAPDSSFWANHDSKYAYWGTSMAPIVAGNVAOLREHFV 273
Db 378 ---DVSAPGSSILSTLNGSTGTTTGS-----ASTASNGTSMASPHVAGVVALVQS--V 425

QY 274 KNRGITPK--PSLLK--AALIAGAADIIGLYPNGNOGWRVTLDRSLNVAAYVNESS-----325
Db 426 APTALTAAVETLLKNTARALPGAC-----SGCGGAGIVNADAATAA--INGSGGGG 477

QY 326 -----SLSTSOKATYSFTATACKPLKISLWSDAPASTTASVTL---VND 367
Db 478 GGGNTLTNGTPTVGLGAATGAELNYTITVPAG-----SGTLTVTTSGGSGD 523

QY 368 LDLVI---TAPNGTOYVGNDFTSYPNDNWDGRNNYVENFINAPQSGTYTIEVOAYNVPVG 424
Db 524 ADLYVRAGSAPTDASAYT-----CRPYRS-----GNAETCTITAP-SCTYYVRLKAYS-----569

QY 425 PQTF 429
Db 570 --TFS 572

RESULT 6
SUBT_BACS9
ID SUBT_BACS9 STANDARD; PRT; 420 AA.
AC P28842;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Subtilisin precursor (EC 3.4.21.62).
GN SUB1.
OS Bacillus sp. (strain TA39).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=29336;
RN [1]

SEQUENCE FROM N.A.
RX MEDLINE-92256481; PubMed=1581352;
Narinx E., Davail S., Feller G., Gerday C.;
RT "Nucleotide and derived amino acid sequence of the subtilisin from
the antarctic psychrotroph *Bacillus TA39*."
RL Biochim. Biophys. Acta 1131:111-113(1992).
CC -!- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE,
IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
for peptide bonds, and a preference for a large uncharged residue
in P1. Hydrolyzes peptide amides.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: STILL ACTIVE AT TEMPERATURES CLOSE TO 0 DEGREES
CELSIUS, IT HAS A MARKED HEAT STABILITY.
CC -!- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF
SPORULATION AND MANY MUTATIONS WHICH BLOCK SPORULATION AT EARLY
STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN
IS NOT NECESSARY FOR NORMAL SPORULATION.
CC -!- SIMILARITY: Belongs to peptidase family S8.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL: X62369; CAA44227.1; -
CC PIR: S23407; S23407.
CC HSP: Q99405; IMPT.
CC MEROPS: S08.00A; -
CC InterPro: IPR000209; Peptidase_S8.
CC Pfam: PF00082; Peptidase_S8; 1.
CC PRINTS: PR00723; SUBTILISIN.
CC PROSITE: PS00136; SUBTILASE_ASP; 1.
CC PROSITE: PS00137; SUBTILASE_HIS; 1.
CC PROSITE: PS00138; SUBTILASE_SER; 1.
KW Hydroxylase; Sporulation; Serine protease; Zymogen; Signal.
FT SIGNAL 1 26
FT PROPEP 27 111
FT CHAIN 112 420
FT ACT_SITE 145 145
FT ACT_SITE 182 182
FT ACT_SITE 360 360
SQ SEQUENCE 420 AA; 44086 MW; AE4F121BD32B26EC CRC64;

Query Match 12.3%; Score 277; DB 1; Length 420;
Best Local Similarity 31.2%; Pred. No. 4.4e-11;
Matches 98; Conservative 44; Mismatches 120; Indels 52; Gaps 16;

QY 21 GGGQIVAVADTGLDTRNDSSMHEAFRGKITAL--YALGR-----NNANDTNGHGTTHVAGS 75
Db 136 GGGINIAVLDTGVNTN-----HPDLRNNEVQCKDFVTGYTYTNNSCTDROGHGTHVAGS 189

QY 76 VLNGSTNK---GMAPOANL-VFQSTMDSGGGLG-GLPSNLQTLFQSYASAGARHTN-S 129
Db 190 ALADGGTNGVYGVPADLWAKYKLGDCSGVADDAIRHAGDOATALTAKVIVNMS 249

QY 130 WGAAYNGAYTTDSRNVDVYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAITVGATENLR 189
Db 250 LGSSGESSLITNAVN---YSYNGKVLIIAAAGNSPGYQGSITGYPCALYNVAVAALEN-K 305

QY 190 PSFGSYADNINHVAFQSSRGPT-KDG-----RIKPDVMAPTETILSARSSLAPDSSFWAN 243
Db 306 VENGTY-----RVADFSSRGYSWTDGDYATOKGDEISAPGAAYST-----W---348

QY 244 HDSKYAYMGCTSMATPIVAGNVAQLRHFHFKNRGITPKPSLLKALITAGAADIIGLYPNG 303
Db 349 FDGGYATISGTSWASPHAAGLAAKIWAQYPSASNVDRGELQYRAY---ENDILSGYVAG 405

QY 304 -----NQGWGRVTL 312
Db [1]

Db 406 YGDDFASGFGFATV 419

RESULT 7

AOL1_THEAQ STANDARD; PRT: 513 AA.

AC P08594;

DT 01-AUG-1988 (Rel. 08, Created)

DT 01-AUG-1990 (Rel. 15, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Aqualysin I precursor (EC 3.4.21.-).

GN PSTI

OS Thermus aquaticus.

OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;

OC Thermus.

OX NCBI_TaxID=271;

RN [1]

RC STRAIN-YTL;

RC MEDLINE=90216674; PubMed=2182621;

RA Terada I., Kwon S.-T., Miyata Y., Matsuzawa H., Ohta T.;

RT "Unique precursor structure of an extracellular protease, aqualysin I, with NH₂- and COOH-terminal pro-sequences and its processing in Escherichia coli.";

RL J. Biol. Chem. 265:6576-6581(1990).

RN [2]

RP SEQUENCE OF 75-442 FROM N.A., AND PARTIAL SEQUENCE.

RC STRAIN-YTL;

RX MEDLINE=88225062; PubMed=3286255;

RA Kwon S.-T., Terada I., Matsuzawa H., Ohta T.;

RT "Nucleotide sequence of the gene for aqualysin I (a thermophilic alkaline serine protease) of Thermus aquaticus YT-1 and characteristics of the deduced primary structure of the enzyme.";

RL Eur. J. Biochem. 173:491-497(1988).

RN [3]

RP SEQUENCE OF 128-170.

RX MEDLINE=88151937; PubMed=3162211;

RA Matsuzawa H., Tokugawa K., Hamaoki M., Mizoguchi M., Taguchi H., Terada I., Kwon S.-T., Ohta T.;

RT "Purification and characterization of aqualysin I (a thermophilic alkaline serine protease) produced by Thermus aquaticus YT-1.";

RL Eur. J. Biochem. 171:441-447(1988).

CC -!- FUNCTION: AQUALYSIN I IS A THERMOPHILIC ALKALINE SERINE PROTEASE. THE OPTIMAL TEMPERATURE FOR ITS CASEINOLYTIC ACTIVITY IS 80 DEGREES CELSIUS.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- DEVELOPMENTAL STAGE: SECRETED FROM THE EARLY STATIONARY PHASE UNTIL THE TIME THE CELLS CEASE TO GROW.

CC -!- PTM: THE N- AND C-TERMINAL PRO-SEQUENCES ARE REMOVED THROUGH THE PROTEOLYTIC ACTIVITY OF AQUALYSIN I ITSELF, IN THAT ORDER. THE C-TERMINAL PRO-SEQUENCE IS REQUIRED FOR TRANSLLOCATION OF THE PROTEASES ACROSS THE OUTER MEMBRANE.

CC -!- PTM: TWO DISULFIDE BONDS ARE PRESENT.

CC -!- SIMILARITY: Belongs to peptidase family S8.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: D90108; BAA14135.1; -

DR EMBL: X07734; CAA30559.1; -

DR PIR: A35742; A35742.

DR HSP: P06873; 2PRK.

DR MEKOPS: S08.051; -

DR InterPro: IPR000209; Peptidase_S8.

DR Pfam: PF00082; Peptidase_S8; 1.

DR PRINTS: PR00723; SUBTILISIN.

DR PROSITE: PS00136; SUBTILASE_ASP; 1.

DR PROSITE: PS00137; SUBTILASE_HIS; 1.

DR PROSITE; PS00138; SUBTILASE_SER; 1.

KW Hydrolase; Serine protease; Zymogen; Signal.

FT SIGNAL 1 14

FT PROPEP 15 127

FT CHAIN 128 408 AQUALYSIN I.

FT PROPEP 409 513

FT ACT_SITE 166 166 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 197 197 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 349 349 CHARGE RELAY SYSTEM (BY SIMILARITY).

SQ SEQUENCE 513 AA; 53913 MW; DDFDFE6D4A50B785 CRC64;

Query Match 11.5%; Score 257.5; DB 1; Length 513;

Best Local Similarity 26.0%; Pred. No. le-09;

Matches 117; Conservative 44; Mismatches 144; Indels 145; Gaps 23;

QY 16 SYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALLY-ALGRITNNANDTNGHGHVAG 74

DB 152 TVTATGGRVNVYVIDTGIRT-----THREGGRRVGYDALG--GNQDCNGHGHVAG 203

QY 75 SVLNGSTNKGMAPOANLVFQSIMD--SGGGLGGLPSNLQTLFQOAYSAGARIHTN---- 128

DB 204 TI---GGVTYGVAKAVNLVAVRVLDGNGSGSTSGVIAGVDWV-----TRNHRPPAVA 252

QY 129 --SWGAAVNGAVYTTDSRNVDDYVRKN---DMTILFAAGNEGPNGGTISAPGTAKNAITVG 183

DB 253 NMSLGGGVSTA-----LDNAVKNSTIAAGVYVAVAAAGNDNANACNYS--PARVAEALTVG 304

QY 184 AT--ENLRPSFGSYADININHAQFSSRGPTKDGRIKPDVMAPTGTFILSARSLAPDSSF 241

DB 305 ATTSDDARASFNSGCV-----DLFAPGASIPSA-----W 335

QY 242 ANHDSKYATMGTSMTATPIVAGNVVAQLREHFVNKRGITP---KPSLLKAALIAAGADIGL 298

DB 336 YTSDTATQTLNGTSMATPHVAG--VAAL--YLEONPSATPASVASAILNGATTGRLSGTGS 392

QY 299 GYPNGQNGWGRVTLDRKSLNAVYVNESSLSSTSKATYSTATACKPLKISLVWSDAPAST 358

DB 393 GSPN-----RLLYSLSSGS-----GSTAPCTS 415

QY 359 ---TASVTLVNDLVLVTPAGNTOY-----VGNDFTSPYNNNDNRN--- 397

DB 416 CSYTGTSLSGGPGDYNF---QPNGTYYSPAGTHRAWLRGPAGTDF-DLYLWRWDGSRWLT 471

QY 398 -----NVENVFINAPOSGTYTIEVOAYN 420

DB 472 VGSSTGPTSEESLSYSGTAGYVLRIVAYS 501

RESULT 8

WPRA_BACSU STANDARD; PRT: 894 AA.

ID WPRA_BACSU

AC P54423; O06726;

DT 01-OCT-1996 (Rel. 34, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DE Cell wall-associated protease precursor (EC 3.4.21.-) [Contains: Cell wall-associated polypeptides CWBP23 and CWBP52].

DE WPRA.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 32-54 AND 414-428.

RC STRAIN=168;

RX MEDLINE=97158234; PubMed=9004506;

RA Margot P., Karamata D.;

RT "The wprA gene of Bacillus subtilis 168, expressed during exponential growth, encodes a cell-wall-associated protease.";

RL Microbiology 142:3437-3444(1996).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=98015415; PubMed=9353931;

Downloaded from <http://ajph.org/> on November 10, 2014

van der Laan J.C., Teplyakov A.V., Kelders H., Kalk K.H., Misset O., Mulleners L.J.M., Dijkstra B.W.;
 "Crystal structure of the high-alkaline serine protease PB92 from Bacillus alcalophilus.";
 Protein Eng. 5:405-411(1992).
 [3]
 RN X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS).
 RX MEDLINE=93078250; PubMed=1447775;
 RA Sobek H., Hecht H.-J., Ahle W., Schomburg D.;
 "X-ray structure determination and comparison of two crystal forms of a variant (Asn15Arg) of the alkaline protease from Bacillus alcalophilus refined at 1.85-A resolution.";
 J. Mol. Biol. 228:108-117(1992).
 [4]
 RN STRUCTURE BY NMR OF 112-380.
 RP STRAIN=PB92;
 RC MEDLINE=97277237; PubMed=9115441;
 RX Martin J.R., Mulder F.A., Karimi-Nejad Y., van der Zwan J., Mariani M., Schipper D., Boelens R.;
 "The solution structure of serine protease PB92 from Bacillus alcalophilus presents a rigid fold with a flexible substrate-binding site.";
 Structure 5:521-532(1997).
 RL -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M65086; AAA22212.1; -;
 DR EMBL; A13738; CAA01128.1; -;
 DR PIR; A49778; A49778.
 DR PDB; 1AH2; 15-APR-98.
 DR MEROPS; S08.038; -;
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE_SP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 DR Hydrolase; Serine protease; zymogen; Signal; 3D-structure.
 FT SIGNAL 1 27
 FT PROPEP 28 112
 FT CHAIN 113 380
 FT ACT_SITE 143 143
 FT ACT_SITE 173 173
 FT ACT_SITE 326 326
 FT HELIX 117 120
 FT TURN 121 123
 FT TURN 125 130
 FT TURN 131 131
 FT TURN 135 136
 FT STRAND 138 142
 FT TURN 150 151
 FT STRAND 156 157
 FT HELIX 173 182
 FT STRAND 198 201
 FT TURN 207 208
 FT HELIX 213 226
 FT STRAND 230 233
 FT HELIX 242 252
 FT TURN 253 255
 FT STRAND 257 258
 FT TURN 267 268
 FT TURN 272 275
 FT STRAND 279 279
 FT STRAND 282 285
 FT STRAND 291 291

FT TURN 299 300
 FT STRAND 303 306
 FT TURN 310 314
 FT TURN 315 317
 FT STRAND 318 322
 FT HELIX 325 342
 FT TURN 344 345
 FT HELIX 348 358
 FT HELIX 375 380
 SQ SEQUENCE 380 AA; 539EA72771B6682C CRC64;
 Query Match 10.7%; Score 240; DB 1; Length 380;
 Best Local Similarity 31.1%; Pred. No. 9.8e-09;
 Matches 91; Conservative 30; Mismatches 90; Indels 82; Gaps 14;
 QY 8 VKADVAQSSYGLYGGQIVAVADTGLDTGRNDSMHEAFRGKITALYALGRINANDNG 67
 DB 122 VQAPAAHNR-GLTSGGVKVAVLDTGIST-----HPDINIRGGASFVGEPE-STODGNG 172
 QY 68 HGTHVAGSV--LGNSTNKGMAPQANLVFQSIMDSGGGLGLPSNLQTLFSQAYSAGARI 125
 DB 173 HGTHVAGTIAALNNSIGVLGVAPNAELYAVKVLGASG--SGSVSSIAQGLEWAGNNGMHV 230
 QY 126 HTNSWGA-----AVNGAYTTDSRNVDVYVRKNDMTILFAAGNEGPNGGTISAPGTA 176
 DB 231 ANLSLGSPPSATLEQAVNSA---TSRGV-----LVVAASGNSG--AGSISYPARY 276
 QY 177 KNATTCATE--NLRPSPGYSADNINHVAFQSSRGPTKDRIKPDVYMAPGTFTLSARSSL 234
 DB 277 ANAMAVGATDNNRASFQVAGL-----DIVAPGVNVQSTVPG- 316
 QY 235 APDSFVWHDKSYAYMGTSMTATPIVAGNVA-----QLREHFVKV 275
 DB 317 -----STVASLNGTSMATPHVAGAAALVKQKPNPSWNVQIRNH-LKN 357
 RESULT 14
 ELYA_BACCS STANDARD; PRT; 380 AA.
 AC P41362;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Alkaline protease precursor (EC 3.4.21.-).
 OS Bacillus clausii.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=79880;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=221 / ATCC 21522 / JCM 9139 / DSM 2512;
 RX MEDLINE=93043753; PubMed=1368952;
 RA Takami H., Kobayashi T., Kobayashi M., Yamamoto M., Nakamura S., Aono R., Horikoshi K.;
 "Molecular cloning, nucleotide sequence, and expression of the structural gene for alkaline serine protease from alkaliphilic Bacillus sp. 221.";
 J. Biol. Chem. 268:1455-1460(1992).
 [2]
 RP SEQUENCE OF 112-129.
 RC STRAIN=221 / ATCC 21522 / JCM 9139 / DSM 2512;
 RA Horikoshi K.;
 RL Microorganisms in alkaline environments, pp.187-194, VCH, Weinheim (1991).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC

Db 120 ANLSLGSPSPSATLEQAVNSA---TSRGV-----LVVAASGNSG--AGSISYPARY 165
Qy 177 KNAITVGATE--NLRFSGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTFFILSARSSL 234
 ||: ||||: | | | | :
Db 166 ANAMAVGATDQNNNRASFQYGAGL-----DIVAPGVNVQSTYPG- 205
Qy 235 APDSSEFWANHDSKYIYMGGTSMATPIVAGNVA-----QLREHFVKN 275
 | | | : ||||| | | | |
Db 206 -----STYASLNGTSMATPHVAGVAALVKOKNPSWSNVQIRNH-LKN 246

Search completed: July 25, 2003, 19:00:05
Job time : 12.0104 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 18:57:15 ; Search time 35.0404 Seconds
(without alignments)
3196.166 Million cell updates/sec

Title: US-09-985-689A-1
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSYGLY.....EQAYNPVGPQTFSIAVN 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2247	100.0	640	Q93UV9	Q93uv9 bacillus sp
2	2183	97.2	639	Q9AQR3	Q9aqr3 bacillus sp
3	2143	95.4	434	Q9AQR0	Q9aqr0 bacillus sp
4	1998.5	88.9	433	Q9AQR1	Q9aqr1 bacillus sp
5	1994.5	88.8	433	Q9AQR4	Q9aqr4 bacillus sp
6	1967.5	88.5	433	Q9AQR2	Q9aqr2 bacillus sp
7	511.5	22.8	1825	Q8T9W1	Q8t9w1 dictyostell
8	447	19.9	1702	Q9GTN7	Q9gtn7 dictyostell
9	414	18.4	654	Q8U0C9	Q8u0c9 pyrococcus
10	406.5	18.1	561	Q8RBJ2	Q8rbj2 thermoaer
11	398	17.7	1239	Q9FBZ4	Q9fbz4 streptomyc
12	381	17.0	430	Q8ENV1	Q8env1 oceanobacil
13	376	16.7	1253	Q9FC06	Q9fc06 streptomyc
14	366	16.3	1102	P95684	P95684 streptomyc
15	349	15.5	1245	Q9RL54	Q9rl54 streptomyc
16	347.5	15.5	444	Q9KBJ7	Q9kbj7 bacillus ha

17	346	15.4	1105	2	Q8KKH6	Q8kkh6 streptomyc
18	338	15.0	412	2	Q9APR6	Q9aer6 thermoanaer
19	338	15.0	1237	2	Q8GGT4	Q8ggt4 streptomyc
20	337	15.0	412	16	Q8RC68	Q8rc68 thermoanaer
21	328.5	14.6	1220	16	Q9LOA0	Q9l0a0 streptomyc
22	327.5	14.6	435	16	Q8EMJ3	Q8emj3 oceanobacil
23	324.5	14.4	824	2	Q45464	Q45464 bacillus sp
24	320.5	14.3	442	16	Q31788	Q31788 bacillus su
25	320.5	14.3	891	1	Q93635	Q93635 thermococcu
26	308.5	13.7	431	2	Q9S3L6	Q9s3l6 bacillus su
27	306.5	13.6	1398	1	Q9P9L1	Q9p9l1 pyrococcus
28	297	13.2	1135	1	Q9P9D1	Q9p9d1 uncultured
29	294.5	13.1	434	2	Q54327	Q54327 bacillus sp
30	282.5	12.6	586	16	Q8PAL8	Q8pal8 xanthomonas
31	281	12.5	419	2	Q45681	Q45681 bacillus su
32	280	12.5	799	16	Q9KEM1	Q9kem1 bacillus ha
33	279.5	12.4	1345	1	Q54437	Q54437 staphyloche
34	278.5	12.4	959	16	Q8PMS7	Q8pms7 xanthomonas
35	277.5	12.3	621	2	Q9F486	Q9f486 alteromonas
36	277.5	12.3	621	2	Q53401	Q53401 alteromonas
37	272.5	12.1	568	16	Q8PMC0	Q8pmc0 xanthomonas
38	272.5	12.1	575	16	Q8PNW1	Q8pnw1 xanthomonas
39	268.5	11.9	403	2	Q45463	Q45463 bacillus sp
40	267	11.9	715	2	P70765	P70765 alteromonas
41	265	11.8	1098	16	Q9L1Z8	Q9l1z8 streptomyc
42	264.5	11.8	966	16	Q8PB28	Q8pb28 xanthomonas
43	263	11.7	1331	17	Q9YEG9	Q9yeg9 aeropyrum p
44	263	11.7	1571	2	Q8GCW3	Q8gcw3 streptococc
45	262.5	11.7	467	2	Q93IQ4	Q93iq4 xanthomonas

ALIGNMENTS

RESULT 1

Q93UV9 ID Q93UV9 PRELIMINARY; PRT; 640 AA.
AC Q93UV9;
DT 01-DEC-2001 (TREMREL. 19, Created)
DT 01-OCT-2002 (TREMREL. 22, Last sequence update)
DT 01-MAR-2003 (TREMREL. 23, Last annotation update)
DE Protease.
DN PROF.
OS Bacillus sp. KSM-KP43.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=109322;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KP43;
RA Itoh S., Saeki K.
RT "new protease"
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB051423; BAB55674.2; -
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
SQ SEQUENCE 640 AA; 67991 MW; 4BBAF7E9D592C15 CRC64;

Query Match 100.0%; Score 2247; DB 2; Length 640;
Best Local Similarity 100.0%; Pred. No. 2, 4e-123;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	NDVARGIVKADVAQSSYGLYCGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN	60
Db	207	NDVARGIVKADVAQSSYGLYCGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN	266
QY	61	NANDTNGHGTTHVAGSVLNGSTNKGMAPQANLVFQSIMDSGGGLGGLPSNLTQFLSQAYS	120
Db	267	NANDTNGHGTTHVAGSVLNGSTNKGMAPQANLVFQSIMDSGGGLGGLPSNLTQFLSQAYS	326
QY	121	AGARIHTNSGAAVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNGTISAPGTAKNAI	180

Db	327	AGARIHNTNSWGAANGVAGTTDSRNVDYVRKNDMTILFAAGNEGNGGTISAPGTAKNAI	386
Qy	181	TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTFILSARSLAPDSSF	240
Db	387	TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTFILSARSLAPDSSF	446
Qy	241	WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLGY	300
Db	447	WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLGY	506
Qy	301	PNGNQGWRVTLDKSLNVAYNVNESSLSSTSQATYSFTATAGPKLISLWSDAPASTTA	360
Db	507	PNGNQGWRVTLDKSLNVAYNVNESSLSSTSQATYSFTATAGPKLISLWSDAPASTTA	566
Qy	361	SVTLVNDLDLVITAPNGTQYVGNDETSPYNDNWDGRNNVENVFINAPQSGTYYTIEVOAYN	420
Db	567	SVTLVNDLDLVITAPNGTQYVGNDETSPYNDNWDGRNNVENVFINAPQSGTYYTIEVOAYN	626
Qy	421	VPVGPQTFSLAIVN 434	
Db	627	VPVGPQTFSLAIVN 640	
RESULT 2			
Qy	Q9AQR3	PRELIMINARY; PRT; 639 AA.	
Ac	Q9AQR3		
Dt	01-JUN-2001 (Tremblrel. 17, Created)		
Dt	01-OCT-2002 (Tremblrel. 22, Last sequence update)		
Dt	01-MAR-2003 (Tremblrel. 23, Last annotation update)		
De	Protease.		
Gn	PROA.		
Os	Bacillus sp. 9860.		
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.		
Ox	NCBI_TaxID=133778;		
Rn	[1]		
Rp	SEQUENCE FROM N.A.		
Rc	STRAIN=9860;		
Rx	MEDLINE=20568675; PubMed=11118284;		
Ra	Saeki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,		
Rt	"Novel oxidatively stable subtilisin-like serine proteases from		
Rt	alkaliphilic Bacillus spp.: enzymatic properties, sequences, and		
Rt	evolutionary relationships."		
Rl	Biochem. Biophys. Res. Commun. 279:313-319(2000).		
Dr	EMBL; AB046403; BAB21266.2; -		
Dr	HSSP; P00782; ISUP.		
Dr	InterPro; IPR000209; Peptidase_S8.		
Dr	Pfam; PF00082; Peptidase_S8; 1.		
Dr	PRINTS; PR00723; SUBTILISIN.		
Dr	PROSITE; PS00137; SUBTILASE_HIS; 1.		
Dr	PROSITE; PS00138; SUBTILASE_SER; 1.		
Sq	SEQUENCE 639 AA; 68185 MW; 316AF6FFDBE4FF54 CRC64;		
Query Match 97.2%; Score 2183; DB 2; Length 639;			
Best Local Similarity 96.3%; Pred. No. 1.3e-119;			
Matches 418; Conservative 13; Mismatches 3; Indels 0; Gaps 0;			
Qy	1	NDVARGIVKADVAGSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITAYALGRTN	60
Db	206	NDVARGIVKADVAGSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITAYALGRTN	265
Qy	61	NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIMDSSGGLGGLPSNLQTLFSQAYS	120
Db	266	NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIMDSSGGLGGLPSNLQTLFSQAYS	325
Qy	121	AGARIHNTNSWGAANGVAGTTDSRNVDYVRKNDMTILFAAGNEGNGGTISAPGTAKNAI	180
Db	326	AGARIHNTNSWGAANGVAGTTDSRNVDYVRKNDMTILFAAGNEGNGGTISAPGTAKNAI	385
Qy	181	TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTFILSARSLAPDSSF	240

Db	386	TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTFILSARSLAPDSSF	445
Qy	241	WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLGY	300
Db	446	WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLGY	505
Qy	301	PNGNQGWRVTLDKSLNVAYNVNESSLSSTSQATYSFTATAGPKLISLWSDAPASTTA	360
Db	506	PNGNQGWRVTLDKSLNVAYNVNESSLSSTSQATYSFTATAGPKLISLWSDAPASTTA	565
Qy	361	SVTLVNDLDLVITAPNGTQYVGNDETSPYNDNWDGRNNVENVFINAPQSGTYYTIEVOAYN	420
Db	566	SVTLVNDLDLVITAPNGTQYVGNDETSPYNDNWDGRNNVENVFINAPQSGTYYTIEVOAYN	625
Qy	421	VPVGPQTFSLAIVN 434	
Db	626	VPVGPQTFSLAIVN 639	
RESULT 3			
Qy	Q9AQR0	PRELIMINARY; PRT; 434 AA.	
Ac	Q9AQR0		
Dt	01-JUN-2001 (Tremblrel. 17, Created)		
Dt	01-JUN-2001 (Tremblrel. 17, Last sequence update)		
Dt	01-OCT-2002 (Tremblrel. 22, Last annotation update)		
De	Protease (Fragment).		
Gn	PROE.		
Os	Bacillus sp. NV1.		
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.		
Ox	NCBI_TaxID=133781;		
Rn	[1]		
Rp	SEQUENCE FROM N.A.		
Rc	STRAIN=NV1.		
Rx	MEDLINE=20568675; PubMed=11118284;		
Ra	Saeki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,		
Rt	"Novel oxidatively stable subtilisin-like serine proteases from		
Rt	alkaliphilic Bacillus spp.: enzymatic properties, sequences, and		
Rt	evolutionary relationships."		
Rl	Biochem. Biophys. Res. Commun. 279:313-319(2000).		
Cc	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.		
Dr	EMBL; AB046406; BAB21269.1; -		
Dr	HSSP; P00782; ISUP.		
Dr	InterPro; IPR000209; Peptidase_S8.		
Dr	Pfam; PF00082; Peptidase_S8; 2.		
Dr	PRINTS; PR00723; SUBTILISIN.		
Dr	PROSITE; PS00137; SUBTILASE_HIS; 1.		
Dr	PROSITE; PS00138; SUBTILASE_SER; 1.		
Kw	Hydrolase; Protease; Serine protease.		
Ft	NON_TER 1		
Ft	NON_TER 434 434		
Sq	SEQUENCE 434 AA; 45294 MW; 83517EDDB74125D2 CRC64;		
Query Match 95.4%; Score 2143; DB 2; Length 434;			
Best Local Similarity 93.5%; Pred. No. 1.7e-117;			
Matches 406; Conservative 19; Mismatches 9; Indels 0; Gaps 0;			
Qy	1	NDVARGIVKADVAGSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITAYALGRTN	60
Db	1	NDVARGIVKADVAGSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITAYALGRTN	60
Qy	61	NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIMDSSGGLGGLPSNLQTLFSQAYS	120
Db	61	NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIMDSSGGLGGLPSNLQTLFSQAYS	120
Qy	121	AGARIHNTNSWGAANGVAGTTDSRNVDYVRKNDMTILFAAGNEGNGGTISAPGTAKNAI	180
Db	121	AGARIHNTNSWGAANGVAGTTDSRNVDYVRKNDMTILFAAGNEGNGGTISAPGTAKNAI	180
Qy	181	TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTFILSARSLAPDSSF	240
Db	181	TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTFILSARSLAPDSSF	240

Qy	241	WANHDSKYAYMGGTSMATPIVAGNVAQALREHFVKNRGITPKPSLLKAALIAGAADIGLCY	300
Db	241	WANHDSKYAYMGGTSMATPIVAGNVAQALREHFVKNRGITPKPSLLKAALIAGATDGLCY	300
Qy	301	PNGNOGWRVLDKSLNVAIYNNESSLSSTSOKATYSFTATAGKPKLISLVWSDAPASTTA	360
Db	301	PSGNOGWRVLDKSLNVAIYNFNETSSLTNOKATYSFTAQSGKPKLISLVWSDAPASTSA	360
Qy	361	SVTLVNDLDLVIATPANGTQYVGNDFTPSYNDNWDGRNNVENVFINAPOSGTYTIEVQAYN	420
Db	361	SVTLVNDLDLVIATPANGTKYVGNDETAPYDNNWDGRNNVENVFINAPOSGTYTIEVQAYN	420
Qy	421	VPVGPQTFSLAIVN	434
Db	421	VPQGPQAFSLAIVN	434

```

RESULT 4
Q9AQR1 PRELIMINARY; PRT; 433 AA.
ID Q9AQR1 AC Q9AQR1;
AC Q9AQR1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Protease (Fragment).
GN PROD.
OS Bacillus sp. S052l.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
NCBI_TaxID=133780;
[1]
RN RN SEQUENCE FROM N.A.
RP RP STRAIN=SD521;
RC RC MEDLINE=20568675; PubMed=11118284;
RA Saeki K.; Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,
RA Horikoshi K.;
RT "Novel oxidatively stable subtilisin-like serine proteases from
RT alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
RT evolutionary relationships."
RL Biochem. Biophys. Res. Commun. 279:313-319(2000).
CC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
DR EMBL: AB046405; BAB21268.1; -.
DR HSSP: Q45670; IDB1.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 2.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Protease; Serine protease.
FT NON_TER 1
FT NON_TER 433
SO SEQUENCE 433 AA. 45576 MW. 98170 RESIDUES
```

[illegible]

```

RESULT 5
Q9AQR4 Q9AQR4 PRELIMINARY; PRT; 433 AA.
AC Q9AQR4:
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2003 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Protease (Fragment).
GN PROA.
OS Bacillus sp. D6.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OC NCBI_TaxID=127889;
[1]
RN RN SEQUENCE FROM N.A.
RP STRAIN=D6;
RC
RX MEDLINE=20568675; PubMed=11118284;
RA Saeiki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,
RA Horikoshi K.;
RT "Novel oxidatively stable subtilisin-like serine proteases from
RT alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
RT evolutionary relationships.";
RL Biochem. Biophys. Res. Commun.. 279:313-319(2000).
CC 1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
DR EMBL: AB046402; BAB21265.1; -.
DR HSSP: Q45670; 1DBI.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 2.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Protease; Serine protease.
FT NON_TER 1
FT FT 433
SQ SEQUENCE 433 AA 45636 MW: 52087E0A2516107F CRC64:

```

300 PNGDQGWRTLDKSLNVAAYVNEATATATGOKATYSFOAQAGKPLKISLVWTDAPGSTTA 359
361 SVTLVNDLDELIVITAPNGTQYVGNDFTSYNDNDGRNNVENFVINAPOSQGTITIEVQAYN 420
360 SYTLVNDLDELIVITAPNGKQYVGNDFSYPDNDNDGRNNVENFVINAPOSQGTITIEVQAYN 419
421 VPVGPQTFSLAIYN 434
420 VPSGPQRFSLAIHV 433

RESULT 7
Q8T9W1 PRELIMINARY; PRT; 1825 AA.
ID Q8T9W1
AC Q8T9W1; 2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Serine protease/ABC transporter TagD.
GN TAGD.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ax4;
RA Anjard C., Loomis W.F.;
RT "Evolution of the ABC transporters of Dictyostelium.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; AF466309; AAL/4253.1; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR001140; ABC_TM_transporter.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00664; ABC_membrane; 1.
DR Pfam; PF00005; ABC_tran; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR ProDom; PD00006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR ATP-binding; Protease; Transport.
KW ATP-binding; Protease; Transport.
SQ SEQUENCE 1825 AA; 202641 MW; E28160BC78613A3B CRC64;

Query Match 22.8%; Score 511.5; DB 5; Length 1825;
Best Local Similarity 27.9%; Pred. No. 2.2e-21;
Matches 164; Conservative 76; Mismatches 155; Indels 193; Gaps 22;

QY 19 LYQGGQIVAVADTGLDGTGR---NDS-----SMHEAFRGKITALYALGRNNANDTNGH 68
DB 327 LRKGQILSTADTGLDGSCHGFFSDSNPNIPYNSVNLNRKVVTVIGSL--HDNEDYVDGH 384
QY 69 GTHVAGSVLNG-----STNKGMAPQANLVFQSI-MDSGGGLGSLNLTQTLFSQAYS 120
DB 385 GTHVCGSAAGAPEDSSIAISSFSGLATDAKIAFPDLADSPNSNEPVPEDYSOLYQPLYN 444
QY 121 AGARIHTNSWGA-----AVNCAYTDSRVDVYVKN-DMTILFAGNEGPNNGTISAPGT 175
DB 445 AGARVHGDSDGWSGLSIQGYLGSYDDAGSIDDFLTHPDIILIRAAAGNEQYSSLLS-QAT 503
QY 176 AKNAITVGTATENLRPSF-----
DB 504 AKNVIIVGAQTTHESYTTDALEYSNFETVAKSTLNSLCQSFDDKCYTVTYTAQCCTEYST 563
QY 193 -----GSYAD-----NINHVAOFSSRGPTKDGRIKPDVMAPGTILSA 230
DB 564 VKGLSGCCTSYIKNSYASIFSSQPELYNENNICSFSKGPETHDGLKPKDIPVAPGQVITSA 623
QY 231 RSSIA-----PDSSEWANHDSKYATMGTSMTATPIVAGNVAQLREH----- 271

301 PNGDQGWRTLDKSLNVAAYVNEATATATGOKATYSFOAQAGKPLKISLVWSDAPASTTA 360
300 PSGDQGWRTLDKSLNVAAYVNEATATGOKATYSFQTAQAGKPLKISLVWTDAPGSTTA 359
361 SVTLVNDLDELIVITAPNGTQYVGNDFTSYNDNDGRNNVENFVINAPOSQGTITIEVQAYN 420
360 SYTLVNDLDELIVITAPNGKQYVGNDFSYPDNDNDGRNNVENFVINAPOSQGTITIEVQAYN 419
421 VPVGPQTFSLAIYN 434
420 VPSGPQRFSLAIHV 433

RESULT 6
Q9AQR2 PRELIMINARY; PRT; 433 AA.
ID Q9AQR2
AC Q9AQR2; 2002 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Protease (Fragment).
GN PRO.
OS Bacillus sp. Y.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=133779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Sasaki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,
RA Horikoshi K.;
RT "Novel oxidatively stable subtilisin-like serine proteases from
RT alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
RT evolutionary relationships.";
RL Biochem. Biophys. Res. Commun. 279:313-319(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
DR EMBL; AB046404; BAB21267.1; -.
DR HSP; O45670; IDBI.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 2.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR Hydrolase; Protease; Serine protease.
FT NON_TER 1 1
FT NON_TER 433 433
SQ SEQUENCE 433 AA; 45587 MW; B81291A803C775AE CRC64;

Query Match 88.5%; Score 1987.5; DB 2; Length 433;
Best Local Similarity 87.3%; Pred. No. 2.1e-108;
Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIKADVAQSSYGLYGQIVAVADTGLDGRNDSMSHEAFRGKITALYALGRTN 60
DB 1 NDVARGIKADVAQNNYGLYGQIVAVADTGLDGRNDSMSHEAFRGKITALYALGRTN 60
QY 61 NANTDNGHGHVAGSVLNGSTNKGMAPQANLVFQSIMDSSGGGLGSLNLTQTLFSQAYS 120
DB 61 NASDPNGHGHVAGSVLNG-ALNKGMAPQANLVFQSIMDSSGGGLGSLNLTQTLFSQAWN 119
QY 121 AGARTHTNSWGAAGVNCAYTTDSRVDVYVKN-DMTILFAGNEGPNNGTISAPGTAKNAI 180
DB 120 AGARLHTNSWGAAPVNGAYTANRQVDEYVRNNDMTVLFAAGNEGPNNGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGYADNINHVAOFSSRGPTKDGRIKPDVMAPGTITLSARSSLAPDSSF 240
DB 180 TVGATENLRPSFGSIADNPNHIAQFSSRGATDGRKIPDVTAGTITLSARSSLAPDSSF 239
QY 241 WANHDSKYATMGTSMTATPIVAGNVAQLRHFVKNRGIPTKPSLLKAALTAGAADIGLGY 300
DB 240 WANYSKYATMGTSMTATPIVAGNVAQLRHFVKNRGIPTKPSLLKAALTAGAADIGLGY 299
QY 301 PNGDQGWRTLDKSLNVAAYVNEATATATGOKATYSFOAQAGKPLKISLVWSDAPASTTA 360

```

Db 624 RSNQANTDQCQGDGSLPNTNALLSE-----SGTSMATPLATAATTILROYLVDCYYPT 676
QY 272 --FYKNRGITPKPSLLKAAIAGAADIGLGP-----NGNQGWGRV 310
Db 677 GSIVESKLQPTGSLKALMINNAQLNGTFPLSSTNTNPSNAVEDTFAGANFVGWGS 736
QY 311 TLDKSLNVAAYNESS-----SLTSOKATYSFT-----338
Db 737 RMSEWL---YVESSGVKPKPSRWVGIGELGDKKASNKKEYSLSTGQNVSYCFYKPPSS 793
QY 339 --ATAGKP-LKLSLWSDAPASTTASVTLVNDLDLVIAPNGTQ-----YVGNDFTSYND 391
Db 794 GNSGGIPRIVATLVMTDPPSYGAKLNLVNLDTMT---NTESEFIFYNSGGSSYNG 850
QY 392 N-----WDGRNNVNFV---INAPQSGTGTIEQAYNPVGPOTFS 429
Db 851 TKGTTPLQDSINNVEGIYTPINTKSEISPRFIAGTNPIGPNFS 898

RESULT 8
Q9GTN7
ID Q9GTN7 PRELIMINARY: PRT: 1702 AA.
AC Q9GTN7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE TAGA.
GN TAGA.
OS Dictyostelium discoideum (slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
SEQUENCE FROM N.A.
RA Good J.R., Cabral M., Kuspa A.;
RT "TagA, a putative serine protease/ABC transporter of Dictyostelium
RT that is expressed at the onset of development and is required for the
RT differentiation of a subpopulation of prespore cells.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL: AF263455; AAG11416.1; -.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR011140; ABC_TM_transpt.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00664; ABC_membrane; 1.
DR Pfam: PF00005; ABC_tran; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR ProDom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
DR ATP-binding; Transport.
KW ATP-binding; Transport.
SQ SEQUENCE 1702 AA; 187103 MW; 4A67716303CB7131 CRC64;

Query Match 19.9%; Score 447; DB 5; Length 1702;
Best Local Similarity 25.8%; Pred. No. 1.2e-17;
Matches 140; Conservative 99; Mismatches 164; Indels 140; Gaps 22;

QY 18 GLYGOGIYAVADTGLDTCR---NDSSMHEAFRGKITALLYALGRTNNDNTNGHGHVAG 74
Db 300 GLKGDEIVGCADTGIDINHCFFYDTPNPGTSTHKKIIS-YSSGNGQDQIDEIDGHGTHVG 358
QY 75 SVLNGSTN-----KGMAPQANLVFQSI-MDSGGGLGLPSNLQTLFQAYSAGARIHT 127
Db 359 TIIGSTTVDPVSSEFSGGAPNSKVAFVDLOVSGNGL-SIQSNLTAIYOSTYDONAKVHC 417
QY 128 NSWGAANGVAYTDSRNVDYVRKN-DMTILFAAGNEGPNNG-----TISAPGTAKNAIT 181
Db 418 DAWNSNIGPFYTGVTMIDRFQWDHDFLWRSAGN-NVNFEGNSIYILSQESTSKSLV 476
QY 182 VCATENLRPSFGSYADNINH-----201
Db 477 VGSSN--OPS-STYLLSSIDWDWDFIYNSIRTSVCTQGOSIYGITCSDVPTQTTSVDIQT 533

```

```

QY 202 -----VAOFSSRGPTKDGRIKPDVMAQPTFILS 229
Db 534 QCCSNPILAKICCCSTEIQOQYQTNSTVSEFIPSLFSGVGPTSDGRKLPDLIAPGSPILS 593
QY 230 ARSSLAPDSSEFWAH-----DSKAYMGCTSMATPIVAGNVAQLRE-----HFVKN 275
Db 594 SR-SLGPSSIT--NHCSPITSGIATSALIAMEGSSQAAAVATSAAVLVHQYVRDGYFING 650
QY 276 R-----GITPKPSLLKAAIAGAA---DIGLGPNGNOGWGRVTLTDKSLNVAAYNESS--- 325
Db 651 KVNSSVGFQPSASLVKATLINTASINVDSTLEY---SOGFGNIQLSKLITTTNAOTTSLD 707
QY 326 -----SLTSOKATYSFTATAGKPLKLSLWSDAPASTTASVTLVNDLDLVIITA- 374
Db 708 IPSIEKADPIINTGETNSYCFSLDSKADIDITLVWTDPAQSPSTFTLVNLDLALLAF 767
QY 375 PNG--TOYVGNDFTSYNDN---WDGRNNVNFVINAPOSQGTIYIEQAYNPVGPOTFS 429
Db 768 VDGELSIYSGNSETIFKNTSOVIFDQLNNVEIRIKAPIGSYDVKIFGTNIVIPNQSYS 827
QY 430 LAI 432
Db 828 VVI 830

RESULT 9
Q8UOC9
ID Q8UOC9 PRELIMINARY: PRT: 654 AA.
AC Q8UOC9;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Alkaline serine protease.
GN PFI670.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcaceae; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE010265; AAL81794.1; -.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
KW Protease; Complete proteome.
SQ SEQUENCE 654 AA; 70230 MW; 1CB145A5F505DB34 CRC64;

Query Match 18.4%; Score 414; DB 17; Length 654;
Best Local Similarity 29.8%; Pred. No. 2.8e-16;
Matches 136; Conservative 59; Mismatches 147; Indels 114; Gaps 19;

QY 18 GLYGOGIYAVADTGLDTCRNDSSMHEAFRGKITALLYALGRTNAN-----DTNGHGH 71
Db 154 GYDGSITIGIITDGTID-----ASHPDLQGV-----IGWVDFVAGRSYPYDDHGHGH 202
QY 72 VAGSLVNGSTN-----KGMAPQANLVFQSI-MDSGGGLGLPSNLQTLFQAYSAGARI 125
Db 203 VASTAAGTGAASNGYKGMPCAKLAGIKVLGADGSGSTIILKGVWAVDNKDKYGIKV 262
QY 126 HTNSWA-----AVNGAYTDSRNVDYVRKNMDMTILFAAGNEGPNNGTISAP 173
Db 263 INLSGSSQSDGTDLASQAVNAWDA-----GLVWVAAGNSGPNKYITIGSP 310
QY 174 GTAKNAITVGATENLRPSFGSYADNINHVAOFSSRGPTKDGRIKPDVMAQPTFILSARS 233
Db 311 AAASKVITVGA-----VDKYDVTTSFSSRGPTADGRKLPVAVAPGNWIIAARAS 359

```


KW Complete proteome.
SQ SEQUENCE 1239 AA; 128505 MW; 8F5E9AC68EB1260A CRC64;
Query Match 17.78; Score 398; DB 16; Length 1239;
Best Local Similarity 31.18; Pred. No. 5.6e-15;
Matches 144; Conservative 55; Mismatches 182; Indels 82; Gaps 16;
QY 8 VKADVAQSSY-----GLYGGQIVAVADTGLDGRNDSSMHEAFRCKITIALYALG 57
Db VEADLADSTAOICAPRAWAGNTGGQVEVAVLDTGVDA-----HPDLADRIAAROSFV 272
QY 58 RINNANDTNGHGHVAGSVLNGSTN----KGMAPQANLVFQSTMDSGGLGLPLPSNLOT 113
Db PDENTDRDCHGHVASTIAGTGAASAGKEGVAPGARLSIGKVLDN--SGRGOISWTLAA 331
QY 114 LFSQAYSGARITHNSGWAAGVNGAYTDSRNVDVYRKNDMTILFRAAGNEPNGGTISA 172
Db MEWAVERIAKIVNMSLGSQSGDSDPMSRAVDRI SAQTGALFVVAAGN--GGEAGSIGA 390
QY 173 PGTAKNATVGTATENLRPSFGSYADNINHVAFSSRGTDKGRIPKDVMAAGTFTLSARS 232
Db PGVATSAITVGA-----VDATDTLAPFSSQGPVVDGALKPEITAPGVGILAA-- 437
QY 233 SLAPDSSFWANHDSKYAYMGCTSMATPIVAGNVAQLREHFEVKNRGITPKPSLKAAL-- 289
Db -----NSSFAAGNGAYQSLSGTSMATPHVAGAALL-----AAARPDLSGSAKDV 484
QY 290 IAGAADIGLPGNGOGWGRVTLDKSLN-----VAYVNESSLSSTQKATYSFTATAG 342
Db LASSSHRTPRYDAFQAGSGRVDVDAVRAGVYASATAYAPGSSGPPVRLVTVNTGTAA 544
QY 343 KPLKLSLWSDA-----PASTTASVTLVNDLVLITAPNGTQYVGNDFTSF 388
Db VTELSVAATHAPEGVFLRSASRVTPAHGTADVTLTIDGS---GSAGGRAYSGQILAT- 600
QY 389 YNDNWDGRNNVENVFNAPOSCTYIEVQ-AYNVPVGPQTFSL 430
Db ---DADARNVAHTAVSAGVPRHKLIVHFKDAGNPV-PGVFDL 639
RESULT 12
Q8ENV1 PRELIMINARY; PRT; 430 AA.
AC Q8ENV1;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Intracellular alkaline serine proteinase.
GN OB2375.
OS Oceanobacillus thelyensis.
OC Bacteria; Firmicutes; Bacillales; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
RX MEDLINE=2220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus thelyensis isolated from the Iheya Ridge and its unexpected adaptive capabilities to extreme environments.";
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL; AP004601; BAC14331.1; -;
KW Complete proteome.
SQ SEQUENCE 430 AA; 45838 MW; 6D09A99B8C1E310F CRC64;
Query Match 17.08; Score 381; DB 16; Length 430;
Best Local Similarity 34.88; Pred. No. 1.4e-14;
Matches 117; Conservative 53; Mismatches 126; Indels 40; Gaps 16;
QY 2 DVARGIVKADVAOSSYGLYGGQIVAVADTGLDGRNDSSMHEAFRCKIT--ALYALGRT 59
Db DTASSSINADVLKES--GLTGOGSTIAVIDTGIHP-----HEDLEGRIGFADFVKQT 172

QY 60 NNANDTNGHGHVAGSVLNGSTN----KGMAPQANLVFOSIMDSGGGLGLPLSNLOTFL 115
Db EYDD--NGHGHVAGSVLNGSTN----KGMAPQANLVFOSIMDSGGGLGLPLSNLOTFL 229
QY 116 SOAYSAGARIHNSGWAAGVNGAYTDSRNVDVYRKNDMTILFRAAGNEPNGGT 169
Db DMCIONQSYNTNLSLSL--GSDATEPAEGDPVNVAVETAWDNGMVVCAAGNSGPGDKT 288
QY 170 ISAPGTAKNAITVGTATENLRPSFGSYADNINHVAFSSRGTDKGRIPKDVMAAGTFTLS 229
Db VCSGSLSPKVIITVGAADDNNTAERS-----DSDVAEFSRGTIDGLTKPULLTPGVDIVS 344
QY 230 ARS--SLAPDSSFWANHDSKYAYMGCTSMATPIVAGNVAQLREHFEVKNRGITPKPSLKA 287
Db LRAPGSFIDKTKNSARVGSNYISLSTSMATPICAGIVAQLLQ---SDSSLT--PNOVKE 399
QY 288 ALIAGAADIGLPGNGOGWGRVTLDKSLNVAIVNE 323
Db KLMEACQDLGQS-PN-VQAGYL---NAANLININE 430
RESULT 13
Q9FC06 PRELIMINARY; PRT; 1253 AA.
AC Q9FC06;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DE 01-MAR-2001 (TremBLrel. 23, Last annotation update)
DE putative secreted peptidase.
GN SC07176 OR SC8A11.04C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Hopwood D.A.;
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939130; CAC01576.1; -;
DR HSSP; Q99405; 1MPT.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_s8.

DR Pfam: PF02225; PA: 1.
 DR InterPro: IPR00082; Peptidase_S8: 1.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PROSITE: PS00840; PA: 1.
 DR PROSITE: PS00136; SUBTILASE_ASP: 1.
 DR PROSITE: PS00137; SUBTILASE_HIS: 1.
 DR PROSITE: PS00138; SUBTILASE_SER: 1.
 KW Complete proteome.
 SQ SEQUENCE 1253 AA; 130971 MW; AA69B417EEFEDB89 CRC64;

Query Match 16.7%; Score 376; DB 16; Length 1253;
 Best Local Similarity 27.6%; Pred. No. 1.1e-13;
 Matches 142; Conservative 58; Mismatches 145; Indels 170; Gaps 22;

Qy 16 SYGLYGQGIIVAVADTGLDTRNDSSMHEAFKITALVALGRTNNDNTNGHGHVAGS 75
 Db 233 SGGNTGEGVGVALDTGVADG-----HPDFAGRIATAATSFVDPQDVTDRNGHGHVAST 286

Qy 76 VLNGSTN-----KGMAPQANLYFQSIMDSGG-----GLGG 106
 Db 287 VAGTGAASGGVEKGVAPGASLHIGKVLNDSGQDSWVLGMEWAVRDQHAQIVMSLGD 346

Qy 107 LPSNLQTLFSQAYSAGARIHTNSWGAAVNGAYTTDSRNVDYVYRKNDDMTILFAAGNNGPN 166
 Db 347 SPTDGTDLSEAV-----NLSAETGA-----LFVVAAGNSGPE 380

Qy 167 GGTISAPGTAKNAITVGATENLRPSFGSYADNINHVAQFSSRGP-TKDGRIKPDVMAPGT 225
 Db 381 AYTGTTPAAADAALIVGAVNG-PGKG-----VDQLADFSRGPVRVGDNAVKPDLTAPGV 433

Qy 226 FILSARSLAPSSFWANHDSKYAYMGTTSMATPIVAGNVAQLREHVEVKNRGITPKPSLL 285
 Db 434 GVLARSRYAEG-----EGAYQSLSGTSMATPHVAGAAALLAAEHPDWTG-----QRL 482

Qy 286 KAALIAGAADIIGLYPENGQ-----GWRVTLDKSLNVAIYVNESSLSSTQ---KATY 335
 Db 483 KEALVCTTA-----GTQRESFPDAGSGRV-----DVAARVSTLLASGDFAQAQHY 528

Qy 336 -----SFTATACKPLKISLVSDA-----PASTTASVTLVND 367
 Db 529 PYTPGQTVRRDVTYITNSGPAPVALDLALSPAELPEGLFTTSEAOVTPAHGHTASGVITH 588

Qy 368 LD-----LVITAPNGTOYVGNDFSPYNDNWDGRNNVNFVINA-----POSG 410
 Db 589 LDAAEDNGAYATRLVAGSADGAVLA---RTPVGVNKEGRR-ATLALTAKDHHDKPLSG 642

Qy 411 T-----YTIEVQA-YNVPVGPQTFSL 430
 Db 643 TWILKDVERNTAPKVYSVDASGRLLDLRLSPSTYSV 677

RESULT 14

P95684
 ID P95684 PRELIMINARY; PRT: 1102 AA.
 AC P95684;
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Subtilisin-like protease.
 OS Streptomyces albobogiseolus.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1887;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S-3253;
 RA MEDLINE=97144528; PubMed=8990295;
 RA Suzuki M., Taguchi S., Yamada S., Kojima S., Miura K., Momose H.;
 RT "A novel member of the subtilisin-like protease family from
 Streptomyces albobogiseolus."
 RL J. Bacteriol. 179:430-438(1997).
 CC -/- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
 DR EMBL: D83672; BAA12040.1; -.

DR HSP: P00782; 2SBT.
 DR MEROPS: S08.069; -.
 DR InterPro: IPR002860; GH_BNR.
 DR InterPro: IPR00209; Peptidase_S8.
 DR Pfam: PF02012; BNR: 2.
 DR Pfam: PF00082; Peptidase_S8: 1.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PROSITE: PS00136; SUBTILASE_ASP: 1.
 DR PROSITE: PS00137; SUBTILASE_HIS: 1.
 DR PROSITE: PS00138; SUBTILASE_SER: 1.
 KW Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 1102 AA; 114128 MW; F9E4AD2590FE59E CRC64;

Query Match 16.3%; Score 366; DB 2; Length 1102;
 Best Local Similarity 31.6%; Pred. No. 3.6e-13;
 Matches 148; Conservative 55; Mismatches 175; Indels 90; Gaps 20;

Qy 3 VAR-----GIVKADVAQS-----SYGLYGQGIIVAVADTGLDTRNDSSMHEAFRG 48
 Db 183 VARVWLDGYRKASLDTSVGQIGTPKAWEGYDGKGVKIAVLDTGVD-----ATHPDLKG 236

Qy 49 KITALVALGRTNNDNTNGHGHVAGSVLNGS-----TNKGMAPQANLYFQSIMDSGGGL 104
 Db 237 QVTASKNFTSAPTTGVDVVGHGHVASTAAAGTGAOSKGTGKGVAPGAKILNGKVLDDAG-- 294

Qy 105 GGLPSNLQTLFSQAYSAGARIHTNSWGAAVNGAYTTDSRNVDYVYRKNDDMTILF--AA 160
 Db 295 FGDDSGILAGMEWAAAGGADIVNMSLG---GMDTPETDPLEAAVDKLSAEKGIILFAIAA 350

Qy 161 GNEGPGNGTISAPGTAKNAITVGATENLRPSFGSYADNINHVAQFSSRGP-TKDGRIKPD 219
 Db 351 GNEGPG--STGSPGSDASALTGVA-----VDDKDLADFSSTGPRLCDGAVKPD 397

Qy 220 VMAPGTFILSARSLAPSSFWANHDSKYAYMGTTSMATPIVAGNVAQLREHVEVKNRGIT 279
 Db 398 LTPAGVDITAAAKGNDIAKEVGEKPGAYMTISGTSMATPHVAGAAALLKQOHP----- 452

Qy 280 KPSSLKAAALIAGAADIGLG-YPNQNGQWGRVTLDKSLNVAIYVNESSLS----- 328
 Db 453 WKYAEKLGALTASTKD---GKYTPFEQSGRQVQVDKAITQTVAIEFVSUSFGVQOQPHAD 509

Qy 329 ---TSOKATYSFTATACKPLKISLVSD-----APAS---TTASVTLVNDLVLITAP-NG 377
 Db 510 DKPVTKLTTRNLCTEDVTLLKLTSTATGPKAAAGPTLGCSTL-----TVPANG 561

Qy 378 TQYVGNDFTSYNDNWDGRNNVNFVINAPOS---GTYTIEVQAYNV 421
 Db 562 TASVDVTADTRLGCAVDGTYSAYVYVATGAGOSVRTAAAVEREVESYNV 609

RESULT 15

Q9RL54
 ID Q9RL54 PRELIMINARY; PRT: 1245 AA.
 AC Q9RL54;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Probable secreted peptidase.
 GN SC00432 OR SCF51A.10.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Murphy L., Harris D.;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
 RC Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

Search completed: July 25, 2003, 19:01:25
Job time : 41.0404 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 18:59:51 ; Search time 21.0242 Seconds
(without alignments)
2451.541 Million cell updates/sec

Title: US-09-985-689A-1
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSYGLY.....EYQAYNPVPGPQTSIAIVN 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA:*

1: /cgn2_6/ptodata/2/pubpa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	2247	100.0	434	11	US-09-985-689A-1	Sequence 1, Appli
2	2191	97.5	434	11	US-09-985-689A-2	Sequence 2, Appli
3	2143	95.4	434	11	US-09-985-689A-6	Sequence 6, Appli
4	2125.5	94.6	433	11	US-09-985-689A-7	Sequence 7, Appli
5	1998.5	88.9	433	11	US-09-985-689A-5	Sequence 5, Appli
6	1994.5	88.8	433	11	US-09-985-689A-3	Sequence 3, Appli
7	1987.5	88.5	433	11	US-09-985-689A-4	Sequence 4, Appli
8	452.5	20.1	659	14	US-10-090-624-12	Sequence 12, Appl
9	414	18.4	412	14	US-10-090-624-1	Sequence 1, Appli
10	414	18.4	522	14	US-10-090-624-4	Sequence 4, Appli
11	414	18.4	654	14	US-10-090-624-16	Sequence 16, Appl
12	366	16.3	1079	15	US-10-112-488-39	Sequence 39, Appl
13	363.5	16.2	1208	15	US-10-156-761-13251	Sequence 13251, A
14	346.5	15.4	1139	15	US-10-156-761-10856	Sequence 10856, A
15	304.5	13.6	1398	14	US-10-090-624-6	Sequence 6, Appli

16	283	12.6	580	11	US-09-927-827-55	Sequence 55, Appl
17	280.5	12.5	595	11	US-09-927-827-59	Sequence 59, Appl
18	276	12.3	418	10	US-09-966-921A-2	Sequence 2, Appli
19	270	12.0	1101	15	US-10-156-761-12934	Sequence 12934, A
20	251	11.2	280	15	US-10-209-812-2	Sequence 2, Appli
21	251	11.2	397	11	US-08-779-334A-5	Sequence 5, Appli
22	247	11.0	271	11	US-09-813-408-2	Sequence 2, Appli
23	246.5	11.0	379	11	US-09-813-408-6	Sequence 6, Appli
24	242.5	10.8	627	11	US-09-927-827-60	Sequence 60, Appl
25	240	10.7	269	8	US-08-322-678-10	Sequence 10, Appl
26	240	10.7	269	9	US-09-837-235-16	Sequence 16, Appl
27	240	10.7	269	9	US-09-060-854B-6	Sequence 6, Appli
28	240	10.7	269	10	US-09-975-139-1	Sequence 1, Appli
29	240	10.7	269	10	US-09-976-414-8	Sequence 8, Appli
30	240	10.7	269	14	US-10-075-907-1	Sequence 1, Appli
31	240	10.7	269	14	US-10-075-895-1	Sequence 1, Appli
32	240	10.7	269	15	US-10-033-325-6	Sequence 6, Appli
33	240	10.7	269	15	US-10-209-812-3	Sequence 3, Appli
34	240	10.7	269	15	US-10-104-693-4	Sequence 4, Appli
35	240	10.7	269	15	US-10-228-572-6	Sequence 6, Appli
36	240	10.7	377	11	US-09-813-408-1	Sequence 1, Appli
37	240	10.7	380	11	US-09-824-893A-261	Sequence 261, App
38	238.5	10.6	275	15	US-10-104-693-3	Sequence 3, Appli
39	238.5	10.6	279	11	US-09-813-408-21	Sequence 21, Appl
40	237.5	10.6	382	15	US-10-033-325-2	Sequence 2, Appli
41	237.5	10.6	382	15	US-10-228-572-2	Sequence 2, Appli
42	236	10.5	266	9	US-09-837-235-18	Sequence 18, Appl
43	236	10.5	269	11	US-09-813-408-4	Sequence 4, Appli
44	236	10.5	380	11	US-09-813-408-19	Sequence 19, Appl
45	235.5	10.5	275	8	US-08-322-678-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-09-985-689A-1
; Sequence 1, Application US/09985689A
; Publication No. US20030022351A1
; GENERAL INFORMATION:
; APPLICANT: HATADA, YUJI
; APPLICANT: OGAWA, AKINORI
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483US0
; CURRENT APPLICATION NUMBER: US/09/985,689A
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-1

Query Match 100.0%; Score 2247; DB 11; Length 434;
Best Local Similarity 100.0%; Pred. No. 3e-191;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
|||||
Db 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
|||||
QY 61 NANDTNGHGTIVAGSVLGNSTNKGMAFPQANLVFQSIMDSGGGLGSLPSNLQTLFQAYS 120

Db 61 NNDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSDSGGLGGLPSNLQTLFSQAYS 120
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Db 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTIFLSARSLAPDSSF 240
Db 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTIFLSARSLAPDSSF 240
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAAGADIGLGY 300
Db 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAAGADIGLGY 300
QY 301 PNGNOGWRVTLDKSLNVAYNVNESSLSSTOKATYSFTATAGPKPLKISLWSDAPASTTA 360
Db 301 PNGNOGWRVTLDKSLNVAYNVNESSLSSTOKATYSFTATAGPKPLKISLWSDAPASTTA 360
QY 361 SVTLVNDLDLVTAPNGTQYVGNDFTSYNDNDGNNVNFVINAPOSQGTYYIEVOAYN 420
Db 361 SVTLVNDLDLVTAPNGTQYVGNDFTSYNDNDGNNVNFVINAPOSQGTYYIEVOAYN 420
QY 421 VPVGPQTFSLAIVN 434
Db 421 VPVGPQTFSLAIVN 434

RESULT 2

US-09-985-689A-2

; Sequence 2, Application US/09985689A

; Publication No. US20030022351A1

; GENERAL INFORMATION:

; APPLICANT: HATADA, YUJI

; APPLICANT: OGAWA, AKINORI

; APPLICANT: KAGEYAMA, YASUSHI

; APPLICANT: SATO, TSUYOSHI

; APPLICANT: ARAKI, HIROYUKI

; APPLICANT: SUMITOMO, NOBUYUKI

; APPLICANT: OKUDA, MITSUYOSHI

; APPLICANT: SAEKI, KATSUHIISA

; TITLE OF INVENTION: Alkaline proteases

; FILE REFERENCE: 215483050

; CURRENT APPLICATION NUMBER: US/09/985,689A

; CURRENT FILING DATE: 2002-07-01

; PRIOR APPLICATION NUMBER: JP P2000-355166

; PRIOR FILING DATE: 2000-11-22

; PRIOR APPLICATION NUMBER: JP P2001-114048

; PRIOR FILING DATE: 2001-04-12

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 2

; LENGTH: 434

; TYPE: PRT

; ORGANISM: Bacillus sp.

US-09-985-689A-2

Query Match 97.5%; Score 2191; DB 11; Length 434;
Best Local Similarity 96.5%; Pred. No. 2.8e-186;
Matches 419; Conservative 13; Mismatches 2; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITAIYALGRTN 60
Db 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITAIYALGRTN 60
QY 61 NNDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSDSGGLGGLPSNLQTLFSQAYS 120
Db 61 NNDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSDSGGLGGLPSNLQTLFSQAYS 120
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Db 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180

QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTIFLSARSLAPDSSF 240
Db 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTIFLSARSLAPDSSF 240
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAAGADIGLGY 300
Db 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAAGADIGLGY 300
QY 301 PNGNOGWRVTLDKSLNVAYNVNESSLSSTOKATYSFTATAGPKPLKISLWSDAPASTTA 360
Db 301 PNGNOGWRVTLDKSLNVAYNVNESSLSSTOKATYSFTATAGPKPLKISLWSDAPASTTA 360
QY 361 SVTLVNDLDLVTAPNGTQYVGNDFTSYNDNDGNNVNFVINAPOSQGTYYIEVOAYN 420
Db 361 SVTLVNDLDLVTAPNGTQYVGNDFTSYNDNDGNNVNFVINAPOSQGTYYIEVOAYN 420
QY 421 VPVGPQTFSLAIVN 434
Db 421 VPVGPQTFSLAIVN 434

RESULT 3

US-09-985-689A-6

; Sequence 6, Application US/09985689A

; Publication No. US20030022351A1

; GENERAL INFORMATION:

; APPLICANT: HATADA, YUJI

; APPLICANT: OGAWA, AKINORI

; APPLICANT: KAGEYAMA, YASUSHI

; APPLICANT: SATO, TSUYOSHI

; APPLICANT: ARAKI, HIROYUKI

; APPLICANT: SUMITOMO, NOBUYUKI

; APPLICANT: OKUDA, MITSUYOSHI

; APPLICANT: SAEKI, KATSUHIISA

; TITLE OF INVENTION: Alkaline proteases

; FILE REFERENCE: 215483050

; CURRENT APPLICATION NUMBER: US/09/985,689A

; CURRENT FILING DATE: 2002-07-01

; PRIOR APPLICATION NUMBER: JP P2000-355166

; PRIOR FILING DATE: 2000-11-22

; PRIOR APPLICATION NUMBER: JP P2001-114048

; PRIOR FILING DATE: 2001-04-12

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 6

; LENGTH: 434

; TYPE: PRT

; ORGANISM: Bacillus sp.

US-09-985-689A-6

Query Match 95.4%; Score 2143; DB 11; Length 434;
Best Local Similarity 93.5%; Pred. No. 5.2e-182;
Matches 406; Conservative 19; Mismatches 9; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITAIYALGRTN 60
Db 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITAIYALGRTN 60
QY 61 NNDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSDSGGLGGLPSNLQTLFSQAYS 120
Db 61 NNDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSDSGGLGGLPSNLQTLFSQAYS 120
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Db 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTIFLSARSLAPDSSF 240
Db 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTIFLSARSLAPDSSF 240
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAAGADIGLGY 300
Db 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAAGADIGLGY 300

QY	301	PNGNQGRVTLDKSLNVAYVNESSLTSOKATYSFTATAGKPLKISLVWSDAPASTTA	360
Db	301	PSGNOGGRVTLDKSLNVAFYNETSSLTNOKATYSFTAQSGKPLKISLVWSDAPASTSA	360
QY	361	SVTLVNDLDLVIITAPNGTQYYCNDFTSPYNDWDGRNVNENFINAPOSIGTVEVQAYN	420
Db	361	SVTLVNDLDLVIITAPNGTKYYGNDFTAPYDNNWDGRNVNENFINAPOSIGTVEVQAYN	420
QY	421	VPVGPQTFSLAIVN	434
Db	421	VPOGPQAFSLAIVN	434

```

RESULT 4
US-09-985-689A-7
: Sequence 7, Application US/09985689A
: Publication No. US20030022351A1
: GENERAL INFORMATION:
: APPLICANT: HATADA, YUJI
: APPLICANT: OGAWA, AKINORI
: APPLICANT: KAGEYAMA, YASUSHI
: APPLICANT: SATO, TSYOSHI
: APPLICANT: ARAKI, HIROYUKI
: APPLICANT: SUMITOMO, NOBUYUKI
: APPLICANT: OKUDA, MITSUYOSHI
: APPLICANT: SAEKI, KATSUHISA
: TITLE OF INVENTION: Alkaline proteases
: FILE REFERENCE: 215483US0
: CURRENT APPLICATION NUMBER: US/09/985,689A
: CURRENT FILING DATE: 2002-07-01
: PRIOR APPLICATION NUMBER: JP P2000-355166
: PRIOR FILING DATE: 2000-11-22
: PRIOR APPLICATION NUMBER: JP P2001-114048
: PRIOR FILING DATE: 2001-04-12
: NUMBER OF SEQ ID NOS: 7
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 7
: LENGTH: 433
: TYPE: PRT
: ORGANISM: Bacillus sp.
US-09-985-689A-7

```

Query Match	94.6%	Score 2125.5	DB 11	Length 433
Best Local Similarity	93.5%	Pred. No. 1.9e-180		
Matches 406	Conservative 19	Mismatches 8	Indels 1	Gaps 1
QY	1	NDVARGTVKADVAQAQSSVGLYGQGGIIVAVADTGLDTGNDSSMHEAFRGKITALYALGRTN	60	
Db	1	NDVARGIVKADVAQNPFGLYGQGGIIVAVADTGLDTGNDSSMHEAFRGKITALYALGRTN	60	
QY	61	NANDTNGHGTHVAGSVLGNSTNKGMAPOANLVFQSTMDSGGGLGGLPSNLQTLFSAQAYS	120	
Db	61	NANDPNGHGTHVAGSVLGN-ATNKGMAPOANLVFQSTMDSGGGLGGLPANTLQTLFSAQAYS	119	
QY	121	AGARIHTNSGGAAYNGAYTTDSNRVDYVYRKNDMTILFAAGNEGPGNGTISAPGTAKNAI	180	
Db	120	AGARIHTNSGAPVNGAYTTDSNRVDYVYRKNDMTILFAAGNEGPGSGTISAPGTAKNAI	179	
QY	181	TVGATENLPRSPGSGYADNTHHVAQFSSRGPTKDGRKPDVMAPGTITLSARSSLAPDSSF	240	
Db	180	TVGATENLPRSPGSGYADNTHHVAQFSSRGPTRDGRKPDVMAPGTVILSARSSLAPDSSF	239	
QY	241	WANHDSKYAYMGGTSMATPIVAGNVAQLREHVFVKNRGITPKPSLLKAALIAGAADIGLGY	300	
Db	240	WANHDSKYAYMGGTSMATPIVAGNVAQLREHVFVKNRGVTPKPSLLKAALIAGAADVGLGF	299	
QY	301	PNGNQGWGRVTTDKSLNVAAYVNESSLSTSQKATYSFTATAGKPLKISLWSDAPASTTA	360	
Db	300	PNGNQGWGRVTTDKSLNVAFAVNETSPLTSQKATYSFTAQAGKPLKISLWSDAPGSTTA	359	
QY	361	SVTLVNDLDLVITAPNGTQYVGNDDTSPYNDWDCGRNNVENVFVINAPOSGTYTIEVOAYN	420	

Tue Jul 29 14:12:42 2003

```

; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483US0
; CURRENT APPLICATION NUMBER: US/09/985,689A
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 4
; LENGTH: 433
; TYPE: PR
; ORGANISM: Bacillus sp.
US-09-985-689A-4

```

```

Query Match      88.8%; Score 1987.5; DB 11; Length 433;
Best Local Similarity 87.3%; Pred. No. 3.4e-168; Indels 1; Gaps 1;
Matches 379; Conservative 29; Mismatches 25;

```

```

QY 1 NDVARGIVKADVAQSSYGLYGQGGIVAVADTGLDTRNDSSMHEAFRGKITIYALGRTN 60
Db 1 NDVARGIVKADVAQNNYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITIYALGRTN 60
QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFOSIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 61 NASDPNGHGHVAGSVLNG-ALNKGMAPOANLVFOSIMDSGGGLGGLPSNLQTLFSQAWN 119
QY 121 AGARIHNSWGAAYNGAYTTDSRNVDYVYKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
Db 120 AGARIHNSWGAAYNGAYTANSRQVDEYVRNNDMTVLEAAGNEGPNNGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAPTGFTLSARSSLAPDSSF 240
Db 180 TVGATENLRPSFGSIADPNHIAQFSSRGATRDGRIKPDVTAAGTFTLSARSSLAPDSSF 239
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLY 300
Db 240 WANTNSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLY 299
QY 301 PNGNQGWGRVTLDKSLNVAAYVNESSLSQKATYSFTATAGPKPLKISLVWSDAPASTTA 360
Db 300 PNGDQGWGRVTLKSLNVAAYVNEATATGOKATYSFQAQAGKPLKISLVWTDAPGSTTA 359
QY 361 SVTLVNDLVLITAPNGTOYVGNDFTSYNDNDGRNNVNFENAPQSGTYTIEVQAYN 420
Db 360 SVTLVNDLVLITAPNGQKYVGNDFSYNDNDGRNNVNFENAPQSGTYTIEVQAYN 419
QY 421 VPGVQPTFSLAIVN 434
Db 420 VPSGQRFSLAIVH 433

```

```

RESULT 8
US-10-090-624-12
; Sequence 12, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06

```

```

; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483US0
; CURRENT APPLICATION NUMBER: US/09/985,689A
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3
; LENGTH: 433
; TYPE: PR
; ORGANISM: Bacillus sp.
US-09-985-689A-3

```

```

Query Match      88.8%; Score 1994.5; DB 11; Length 433;
Best Local Similarity 87.6%; Pred. No. 8e-169; Indels 1; Gaps 1;
Matches 380; Conservative 28; Mismatches 25;

```

```

QY 1 NDVARGIVKADVAQSSYGLYGQGGIVAVADTGLDTRNDSSMHEAFRGKITIYALGRTN 60
Db 1 NDVARGIVKADVAQNNYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITIYALGRTN 60
QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFOSIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 61 NASDPNGHGHVAGSVLNG-ALNKGMAPOANLVFOSIMDSGGGLGGLPSNLQTLFSQAWN 119
QY 121 AGARIHNSWGAAYNGAYTTDSRNVDYVYKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
Db 120 AGARIHNSWGAAYNGAYTANSRQVDEYVRNNDMTVLEAAGNEGPNNGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAPTGFTLSARSSLAPDSSF 240
Db 180 TVGATENLRPSFGSIADPNHIAQFSSRGATRDGRIKPDVTAAGTFTLSARSSLAPDSSF 239
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLY 300
Db 240 WANTNSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLY 299
QY 301 PNGNQGWGRVTLDKSLNVAAYVNESSLSQKATYSFTATAGPKPLKISLVWSDAPASTTA 360
Db 300 PNGDQGWGRVTLKSLNVAAYVNEATATGOKATYSFQAQAGKPLKISLVWTDAPGSTTA 359
QY 361 SVTLVNDLVLITAPNGTOYVGNDFTSYNDNDGRNNVNFENAPQSGTYTIEVQAYN 420
Db 360 SVTLVNDLVLITAPNGQKYVGNDFSYNDNDGRNNVNFENAPQSGTYTIEVQAYN 419
QY 421 VPGVQPTFSLAIVN 434
Db 420 VPSGQRFSLAIVH 433

```

```

RESULT 7
US-09-985-689A-4
; Sequence 4, Application US/09985689A
; Publication No. US20030022351A1
; GENERAL INFORMATION:
; APPLICANT: HATADA, YUJI
; APPLICANT: OGAWA, AKINORI

```


Db 375 LTAPGVDTAASAKGNDAIEKVEKPGAGYWTISGTSMAHPVAGAAALIKQOHP----- 429

QY 280 KPSSLKKAALIAAGADIGL-YPNGNQGGRVTLDKSLNVAAYNNESSLS----- 328

Db 430 WKYAEKGAALTASTKD---GKTYPFQSGRGVQDKAITQTVIAEPVLSFGVQOWPHAD 486

QY 329 ---TSOKATYSFTAGKPKLSLWSD-----APAS---TTASVTLVNDLVLITAP-NG 377

Db 487 DKPVTKLTYRNLGFTVTLKLTSTATGPKGAAPAGFTTLGASTL-----TVPANG 538

QY 378 TQYVGNDFTSYNDWNGRNNVFNAPQS-----GTYTIEVOAYNV 421

Db 539 TASVDVTADTRLGGAVDGTYSAIYVATGAGQSVRTAAAREVESYV 586

RESULT 13

US-10-156-761-13251

: Sequence 13251, Application US/10156761

: Publication No. US20030119018A1

: GENERAL INFORMATION:

: APPLICANT: OMURA, SATOSHI

: APPLICANT: IKEDA, HARUO

: APPLICANT: ISHIKAWA, JUN

: APPLICANT: HORIKAWA, HIROSHI

: APPLICANT: SHIBA, TADAYOSHI

: APPLICANT: SAKAKI, YOSHIYUKI

: APPLICANT: HATTORI, MASAHIRA

: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

: FILE REFERENCE: 249-262

: CURRENT APPLICATION NUMBER: US/10/156,761

: CURRENT FILING DATE: 2002-05-29

: PRIOR APPLICATION NUMBER: JP 2001-204089

: PRIOR FILING DATE: 2001-05-30

: PRIOR APPLICATION NUMBER: JP 2001-272697

: PRIOR FILING DATE: 2001-08-02

: NUMBER OF SEQ ID NOS: 15109

: SEQ ID NO 13251

: LENGTH: 1208

: TYPE: PRT

: ORGANISM: Streptomyces avermitilis

US-10-156-761-13251

Query Match 16.2%; Score 363.5; DB 15; Length 1208;

Best Local Similarity 32.0%; Pred. No. 2.8e-23;

Matches 128; Conservative 54; Mismatches 145; Indels 73; Gaps 18;

QY 8 VRADVAQSS-----YGLYGGQIVAVADTGLDTRNDSSMHEAFRGKITALYALG 57

Db 186 VEADMAESNAQIGTRAAMDAGLTGDPVAVLDTGVDT-----THPDLAGVRSRSKFI 239

QY 58 RTNNANDTNGHGTGHVAGSVLNGS-----TNKGMAPOANLVFQSTM-DGGGLGLPSNLQ 112

Db 240 DGEVADRNGHGHVHTVSTVGGSGAASDGTGERGVAPGATLAVGKVLSDQAG-----SESQ 294

QY 113 TLFQSAISA---GARLHTNSWGA--AVNGAYTDD--SRNVDDYVRKNDMTILFAAGNEGP 165

Db 295 IIAGMEWAARDVRAIVSMLSGSTEASDG---TDPMAEAVDTLSEETGALFVVAAGNTGA 351

QY 166 NGGTSAPOTAKNAITVGTATENLRPSFGSYADNINHVAQFSRGPVK-DGRKIPDVMAFG 224

Db 352 -PSSIGSPGAADSALTVA---VDSDDRAAYFTSAGPRHGONALKPDLAAPG 399

QY 225 FTILSARSSLAPDSFWANHDSKYAYMGTSMAATPVAGNVAOLREHFVKNRGTITPKPSL 284

Db 400 VDIAARSQLAPGTG-----YTSMSGTSMAATPVAGNVAOLREHFVKNRGTITPKPSL 452

QY 285 LKAA--LIAGAADIGLPGNQGNGW-RVTLQSLNVA---VNESSLSLTSQ 331

Db 453 MSTSEQLDASVYQLGAGRVSVDPDVGARVATGSDADLGFHRPHDADRPVTKVTYSNSS 512

QY 332 KATYSFT-ATAGKPKLSLWSDA---PASTTASVTLVND 367

Db 513 DTTVELSLAVRGAPAGVATLADTALTVPAHGTAATVTGDD 552

RESULT 14

US-10-156-761-10856

: Sequence 10856, Application US/10156761

: Publication No. US20030119018A1

: GENERAL INFORMATION:

: APPLICANT: OMURA, SATOSHI

: APPLICANT: IKEDA, HARUO

: APPLICANT: ISHIKAWA, JUN

: APPLICANT: HORIKAWA, HIROSHI

: APPLICANT: SHIBA, TADAYOSHI

: APPLICANT: SAKAKI, YOSHIYUKI

: APPLICANT: HATTORI, MASAHIRA

: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

: FILE REFERENCE: 249-262

: CURRENT APPLICATION NUMBER: US/10/156,761

: CURRENT FILING DATE: 2002-05-29

: PRIOR APPLICATION NUMBER: JP 2001-204089

: PRIOR FILING DATE: 2001-05-30

: PRIOR APPLICATION NUMBER: JP 2001-272697

: PRIOR FILING DATE: 2001-08-02

: NUMBER OF SEQ ID NOS: 15109

: SEQ ID NO 10856

: LENGTH: 1139

: TYPE: PRT

: ORGANISM: Streptomyces avermitilis

US-10-156-761-10856

Query Match 15.4%; Score 346.5; DB 15; Length 1139;

Best Local Similarity 29.0%; Pred. No. 8.2e-22;

Matches 143; Conservative 54; Mismatches 179; Indels 117; Gaps 20;

QY 18 GLYGGQIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTNNANDTNGHGTGHVAGSVL 77

Db 249 GYDGGKCVKIAVLDTGVD-----ATHPDLKDQVAESKNFSAAADAAHFGHGTGHVASTAA 302

QY 78 GNGSTN---KGMAPQANLVFQSIMDSGGGLGLPSNLQTLFQSAYSAGARIHTNSWGA 133

Db 303 GTGAKSNGKYKGVAPGATILNGKVLDDTG---SGDDSGILAGMEWAEEQGDVNVLSLG-- 358

QY 134 VNGAYTDDSRNVDDYVRK--NDMTILF--AAGNEGPN--GTISAPGTAKNAITVGTATENL 188

Db 359 --GGDTPEDPLEAEVKNLSEKGIILFAAAGNEGEFGEOTIGSPGSAADALTVGA---- 412

QY 189 RPSFGSYADNINHVAQFSRGPVKDGRKIPDVMAPGTFFILSARSSLAPDSSEFWANHDSK- 247

Db 413 -----VNDSDKLASFSSRGPGLDCAIKPDVTPAGVDITAA---AAPGSVIDQEVGQKP 462

QY 248 --YAYMGTSMAATPVAGNVAOLREHFVKNRGTITPKPSL--LKAALIAGAADIGLPGN 303

Db 463 DGYLTIGTSMATPHVAGAAAILKQOH-----PNWSFAELKGALTGSAR--GGKYTF 513

QY 304 NOGWRVTLDKSLNVAAYNNESSLS-----TSOKATY-----SFTAT 340

Db 514 QQSGRIADVRAIKQSVIANPNSVSPGIIQOWPHITDDKPYTQOLTYRNLGTSVTLNLAST 573

QY 341 AKGP-----LKISLWSDAPASTTASVTLVNDL--VITAPNGTQYV 381

Db 574 ATNPKGVAAPSGFFKLGATKVTVPAGGKASVDFTVNTKLGTTDGTGAYSAYVTATGGQTV 633

QY 382 -----GNDFTSYNDWNGRNNV-----ENVFIPAQSGTGTIEV 416

Db 634 RTAAAVQREVESYDVLKHLDR-DGKPAVNYSTDLTGVSGLAADKWFAPYDASGTVKVRV 692

QY 417 QAYNVPGPQTF 429

Db 693 PKGNFILNASLFA 705

RESULT 15

US-10-090-624-6

: Sequence 6, Application US/10090624

Search completed: July 25, 2003, 19:03:28
Job time : 23.0242 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 28, 2003, 03:38:50 ; Search time 2901.34 seconds
(without alignments)
3635.607 Million cell updates/sec

Title: US-09-985-689A-1
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAYNPVGPQTFLATVN 434

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlp
-Q/cgn2_1/USPTO.spool/US09985689/runat_25072003_143032_845/app_query.fasta_1.1166
-DB-EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cd1 -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09985689 @Cn1_1_5436@runat_25072003_143032_845 -NGPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba.*
2: em_esthum.*
3: em_estlin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_estl.*
10: gb_est2.*
11: gb_htc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	246	10.9	1605	13	BQ622771 CC.Contig
2	241	10.7	640	12	BJ395336 BJ395336
3	225.5	10.0	4198	11	AK029048 Mus muscu
4	224.5	10.0	594	12	BJ393752 BJ393752
5	218.5	9.7	532	29	TA319G10P
6	217	9.7	537	14	CD039158
C 7	207.5	9.2	2141	13	BQ142519
C 8	200	8.9	508	28	AQ652212
9	199.5	8.9	601	12	BJ387574
10	195.5	8.7	1002	29	CNS06D6B
11	195	8.7	771	14	CA320325
12	191	8.5	633	12	BJ369190
13	190	8.5	718	12	BI750157
14	185.5	8.3	574	29	TA315H10P
15	182.5	8.1	614	9	AJ273402
16	180	8.0	641	12	BJ393925
17	180	8.0	665	13	BQ770462
18	179.5	8.0	3091	11	BC011275
19	177.5	7.9	650	9	AJ274038
C 20	176	7.8	530	29	CNS010PO
C 21	176	7.8	2121	29	BZ424995
22	172.5	7.7	580	9	AJ273745
23	172.5	7.7	583	9	AJ273947
24	172.5	7.7	593	9	AJ273918
25	172.5	7.7	601	9	AJ273921
26	171.5	7.6	573	14	CA937626
27	170.5	7.6	601	9	AJ273050
C 28	170.5	7.6	1029	29	CNS071DW
29	170	7.6	895	13	BQ216158
30	170	7.6	1572	11	AY107161
31	169.5	7.5	545	13	BU575479
32	169.5	7.5	604	9	AJ273185
33	168.5	7.5	601	12	BJ365857
34	168.5	7.5	609	9	AJ272712
35	168.5	7.5	937	10	BG246418
36	168	7.5	1050	29	CNS0780L
C 37	167.5	7.5	564	28	AQ651427
38	167.5	7.5	610	9	AJ274218
39	167	7.4	449	12	BJ359939
40	167	7.4	831	14	CA510555
41	164	7.3	615	9	AJ274059
42	164	7.3	781	13	BQ612128
C 43	163	7.3	586	29	CPA561690
44	163	7.3	593	9	AJ273903
45	163	7.3	594	14	CD040823

ALIGNMENTS

RESULT 1
BO622771
LOCUS
DEFINITION
BO622771
ACCESSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE

BO622771
CC.Contig67
cDNA, mRNA sequence.

1605 bp
mRNA
linear
EST 01-JUL-2002
Conidiobolus coronatus
ARSEF 512 Conidiobolus coronatus

BO622771.1
GI:21649940
EST:
Conidiobolus coronatus (Delacroixia coronata)
Conidiobolus coronatus
Eukaryota; Fungi; Zygomycota; Zygomycetes; Entomophthorales;
Ancylistaceae; Conidiobolus.
1 (bases 1 to 1605)


```

/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days neonate"
d17_3575
/note="unnamed protein product: membrane-bound
transcription factor protease, site 1 (MGD|GI:1927235,
GB|NM_019709, evidence: BLASTN, 99%, match=3782)
putative"
/codon_start=1
/protein_id="BAC26263.1"
/db_xref="GI:26325018"
/translation="MKLVSWLLVVLVCGRHLGDRLTGRALKAPCPSCSHLTK
VFPSSTVEYEVAFNGYPTAKARNSFISALLKSEVENWRIIPRNNPSSDPSPDFE
VIQIKERQAGLLTLEDHPNIKRTPQRKVPSRLKFAESNPVPCNETRWSOKWOSSR
PKLRASLSGSFWHATGRHSSRLRAIPROVAOTLQADRLQWQGYTGANTGARVAVFD
TLGSRKHPEKNYKERTNWNTERTLDGLGTFTVAGVTIASMRECQGFAPDAELHIFR
VFTNNOVSYTSWFLDANFIALKKMDVLNLIGGPDFMDHFVKWELWTANNIVWS
AIGNDGPLYTCPLPADQMVDYIGGVIDFEDNIARFSRRGMTTWELPGGYGRKVPDI
TYCAGVGSGVGGCKRLDLRAYOILSSYKPOASLSPSIDLTECPVMWPYCSDPIYVG
RLUPVNMTFGHGKGLDLLRAYOIKPEWRPYLPONGDNIEVAFYSVSLPWSPGYLAISI
GMPTIENVNITLGMVTRGVDKPEWRPYLPONGDNIEVAFYSVSLPWSPGYLAISI
SVTKKAASWEGIAAGHMIIIVASPAETELHSGAEHTSVKLPIKYKLIITPPRSKVL
WDQYNHLRYPPGYPRNLRMNDPLDWNGHDHVRNFDMIOHLRSMGFIHVLEVGAFP
TCDGATGYTILLVDSBEYEPEETAKIRLDVDNGLSLVIFSDWNTSVNRKVKFFDE
NTQWMPMDQTGAANIIPALNELLSVNMNGFSDGLYGEFVLANHDMYIYASGCSTAKFPE
DGVVLTOTFKDOGLEVLQTAFTAVENPVILGLYQIPSEGGRIVLYGDSNCLDDSHRQ
KDCFLLDALLOYTSYGVTTPSLSGNRORPPSAGLAPPAREMECNHLHRYSKVLHA
HLGDPKRPPLACPCHLSNAKEQPLNETAPSNLWKHKLLSIDLKDVLPNFRSNRPOV
RPLSPGESGAWMDIPGGIMPMGRINQEVGQTIPTFAFLGAMVALAFFVVQVISKASRPKR
RRPRAKRPQLAAQHAPRTPSV"
polyA_signal 4180..4185
/note="putative"
polyA_site 4198
/note="putative"
BASE COUNT 1016 a 1065 c 1138 g 979 t
ORIGIN

Alignment Scores:
Pred. No.:          3.85e-12      Length:         4198
Score:             225.50        Matches:         116
Percent Similarity: 41.32%       Conservative:     65
Best Local Similarity: 26.48%    Mismatches:      160
Query Match:       10.04%       Indels:          97
DB:                11           Gaps:            25

US-09-985-689A-1 (1-434) x AK029048 (1-4198)
Qy      8 VallysAlaAspValalacInSerSerTyrglyLeutyrglyGlnGlyAla 27
         :::::::::::::::::::: |||| | |||
Db      1005 CTGACGGCAGATGCTGCTGG---CACATGGGATACACAGGTGCTAATGTGCAGAGTTGCT 1061
         :::::::::::::::::::: |||| | |||

Qy      28 ValAlaAspThrGlyLeuAsnThrGlyArgAsnAspSerMetHisGluAlaPhearg 47
         :::::::::::::::::::: |||| | |||
Db      1062 GTTTTTGTACTCGGGCTC-----AGTCAGAAGCATCCGCATTTTAAG 1103
         :::::::::::::::::::: |||| | |||

Qy      48 GlyLysIleThAlaLeutyrglyAlaLeuGlyArgThrAsnAsnAlaAsn----- 63
         :::::::::::::::::::: |||| | |||
Db      1104 AAT-----GTGAGGAGCAGAACCACTGGACCATTGACGGGCCCTG 1145
         :::::::::::::::::::: |||| | |||

Qy      64 ---AspThrAsnGlyHisGlyThrHisValalaglySerValLeuGlyAsnGlySerThr 82
         ||| :::::::::::::::::::: |||| | |||
Db      1146 GATGATGGCGTAGGCCATGCACATTCGTTGCAGGT---GTGATTGCCAGCATGAGGAG 1202
         :::::::::::::::::::: |||| | |||

Qy      83 AsnLysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerGly 101
         ::||| :::::::::::::::::::: |||| | |||
Db      1203 TGCCCAAGGATTTGCTCCAGATGCAGACGCTGCACATCTTCAGGGCTTTTACCACAAT--- 1259
         ::||| :::::::::::::::::::: |||| | |||

Qy      102 GlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrrSerAla 121
         ::||| :::::::::::::::::::: |||| | |||
Db      1260 -----CAGGTGCTCTACACATCTTGGTTCTTCGGATGCCCTCAACTAT 1301
         ::||| :::::::::::::::::::: |||| | |||

Qy      122 GlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrrThrAsp 141
         ::||| :::::::::::::::::::: |||| | |||
Db      1302 GCCATCCTAAAGAAGATGGAGGTCTCAACCTTTAGCATCGGTGGGCCGCGGATTCATGGAT 1361
         ::||| :::::::::::::::::::: |||| | |||

```

AUTHORS TITLE JOURNAL COMMENT

Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
Full length cDNA of Dictyostellium discoideum at the slug stage
Unpublished
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-811-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES

source
1..594
/organism="Dictyostellium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dds32bl6"
/sex="mat. A"
/dev_stage="Slug stage"
/clone_lib="Dictyostellium discoideum cDNA library, SF"

BASE COUNT 185 a 108 c 107 g 194 t
ORIGIN

Alignment Scores:

Pred. No.: 2,6e-13 Length: 594
Score: 224.50 Matches: 59
Percent Similarity: 50.80% Conservative: 36
Best Local Similarity: 31.55% Mismatches: 59
Query Match: 9.99% Indels: 33
DB: 12 Gaps: 9

US-09-985-689A-1 (1-434) x BJ393752 (1-594)

QY 49 LysIleThrAlaLeuTyAlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsnGlyHis 68
Db 21 AAAGTTTAACTTATTATACCATCAACAGCAGCATAGTATAAAGTGGTGCAC 80
QY 69 GlyThrHisValAlaGlySerValLeuGly-----AsnGlySer 81
Db 81 GTTACACATATTTGTGGTCTGCGAGCAGGTACTCCAGAGGATTTCTTCAGTTAAATATTCA 140
QY 82 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly 101
Db 141 TCATTATAGTGTCTGCAACTGATGCAAGATTCATTC-----TTTGATTGGCA 191
QY 102 GlyGlyLeuGlyGlyLeu-----ProSerAsnLeuGlnThrLeuPheSerGlnAlaTyr 119
Db 192 AGTGGTTTCATCAAGTTTGACACCTCCATCGGATTTGAACAATATATCAACCATTTATAT 251
QY 120 SerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAla-----AlaValAsn 135
Db 252 GACGCAGGTGCAAGATGCGATTCTGATCTTGGGTTCTGTATCAGTAGAGGGGTATACA 311
QY 136 GlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsn---AspMet 154
Db 312 GGTAGTTATTCATCAACACTGCTTCAATGTATGATTTCTCTTTCATCTCATCCAGATTTC 371
QY 155 ThrIleLeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIleSer-----Ala 172
Db 372 ATCATCTTAGAGCTGCTGTRAC-----AACGACCAATACCTATCACTACTCACT 422
QY 173 ProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg----- 189
Db 423 CAATCCACTGCAAGAATTTATTACCGTTGGTGTCTCATCAACAATTCATGAAATTTAT 482
QY 190 -----ProSerPheGlySerTyrAlaAspAsnIleAsn-----His 201
Db 483 TTAAGTATGATGTCCTCAAAATATATAAATTAATCAATCATCTGTCGATATAAATCAAGACTTA 542
QY 202 ValAlaGlnPheSerSerArg 208
Db 543 ATATGTGATTTCGATAGCAGA 563

RESULT 5

TA319G10P
LOCUS
DEFINITION
T. brucei sheared genomic DNA clone 319g10, forward sequence,
genomic survey sequence.
AL492464
VERSION
KEYWORDS
SOURCE
ORGANISM
Trypanosoma brucei
Trypanosoma brucei -
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 532)
Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajandream, M.A. and Barrell, B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nhl@sanger.ac.uk
Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUFat 10.1) was mechanically sheared
to give a tight size distribution (4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES

source
1..532
Location/Qualifiers
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="319g10"

BASE COUNT 112 a 126 c 150 g 143 t 1 others
ORIGIN

Alignment Scores:

Pred. No.: 9.4e-13 Length: 532
Score: 218.50 Matches: 61
Percent Similarity: 48.37% Conservative: 28
Best Local Similarity: 33.15% Mismatches: 68
Query Match: 9.72% Indels: 27
DB: 29 Gaps: 6

US-09-985-689A-1 (1-434) x TA319G10P (1-532)

QY 93 ValPheGlnSerIleMetAspSerGlyGlyGlyLeuGly----- 105
Db 2 GTAATGCCAAGTATTAATGTGTCCGAGGGGCGAAGATTCTTCAGGGGTTCGGGTGCC 61
QY 106 -----GlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyr 119
Db 62 CATCCCATGACAGCTTGTCTCCGCCACGACGCTTACTCAAAATTATTCGTCGGGTATAT 121
QY 120 SerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 139
Db 122 GCGCTGAGCCCGGTGTCTCTCAAACTCGTGGGTTTGTGTCTCCCTCCGAGTATTC 181
QY 140 ThrAspSerArgAsnValAspTyrValArgLys---AsnAspMetThrIleLeuPhe 158
Db 182 GCTGTGGAAGAGATATGATGATGAGTTTCGAGTAGTTATGACGATCGCTACTTATCTTC 241
QY 159 AlaAlaGlyAsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsn 178
Db 242 TCCACTGGCAACAGTATCCAAAGTGGC-----CTAATGACTCCGTGCTGTTGAAGAAC 295

QY	179	AlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsn	198
Db	296	GTGATGGCGTGGGTGCACACAAACAGTG-----TTTGACGCTTCGAAAGAC----	343
QY	199	IleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysPro	218
Db	344	-----ATTGTTTCTCTGTTTCTTCGTCATGCTCAACATACAGCGGTAGGATGAACCC	397
QY	219	AspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSer	238
Db	398	GATCTTGTGGTCCCGGGGAAGAGGTGCTGCTCTTCTCTGCGCAAGCATCAGCT	457
QY	239	SerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThr	258
Db	458	AAA-----CAATGTAAGTGGTGGCGCAAGCGGNGTTCATCGATGGCAACT	502
QY	259	ProIleValAla 262	
Db	503	CGGCGCGTGGCG 514	
RESULT 6			
CD039158		537 bp mRNA linear EST 09-MAY-2003	
LOCUS		psHA005iG10r.206350 psHA: Soybean host 48 hrs post infection	
DEFINITION		Phytophthora sojae cDNA clone SHA005G10 5, mRNA sequence.	
ACCESSION		CD039158	
VERSION		1	
KEYWORDS		Phytophthora sojae	
SOURCE		Phytophthora sojae	
ORGANISM		Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae; Phytophthora	
REFERENCE		1 (bases 1 to 537)	
AUTHORS		Qutob,D., Hraber,P.T., Sobral,B.W.S. and Gijzen,M.	
TITLE		Comparative analysis of expressed sequences in Phytophthora sojae	
JOURNAL		Plant Physiol. 123 (1), 243-254 (2000)	
MEDLINE		20267956	
PUBMED		10806241	
COMMENT		Contact: Tyler B Tyler lab	
FEATURES			
source			
BASE COUNT		102 a 174 c 155 g 104 t	
ORIGIN			
Alignment Scores:			
Pred. No.:		1.37e-12	537
Score:		217.00	69
Percent Similarity:		48.07%	18
Best Local Similarity:		38.12%	48
Query Match:		9.66%	46
DB:		14	8
US-09-985-689a-1 (1-434) x CD039158 (1-537)			
QY	169	ThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeu	188
Db	9	ACGATCAGCAGTCGTCAGGAGCAAGAACGTCATCTCAGTCGGCGCCAGTCTGAACGCC	68
QY	189	ArgProSerPheGlySer-----TyrAlaAspAsnIleAsnHisValAlaGln	204
Db	69	CGGCTTCATTCCTCTCGATCCATCCAGTCGGTCTGCTGATGAGAACACGGTGGCTCG	128
QY	205	PheSerSerArgGlyProThrLysAspGlyArgIleLysProAspValMetAlaProGly	224
Db	129	TTCTCTCCATTTGGACCGACGCTCGACGACGACAGAGCCGACATCGTAGCCCCGGC	188
QY	225	ThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHis	244
Db	189	ATGTCCATCACCTCTTCGACGATCCGAG-----AACCTGGATCGAC-----ACCAAG	236
QY	245	AspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsn	264
Db	237	TCGTCGGCGGTCTGCTCGTCGAGGAACTTCCAGGCTACCCCGTGGTGGCGGTATG	296
QY	265	ValAla-----GlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro	280
Db	297	GCTGTGCTCATCTACGAGTGGTGGTGGTGGTGGAGAAC-----GGCGTG---CCG	350
QY	281	LysPro-----SerLeuLeuLysAlaAlaLeuIleAla	291
Db	351	GACCCACCTACGGCATGATAGATCCCGGCGTCTCTGATTAAAGGCACTTCTTCTACAC	410
QY	292	Gly-----	292
Db	411	ACTGGCGAGGCGATGTCGCCGTCGTGATCGAGCCCGACACTGGCGTTACGTCGCGTG	470
QY	293	-----AlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGly	308
Db	471	GGCGTGGAGACTGCTGCTGAAGCGCTNAACTCTGATCCCGCACTTCAACACGAGGCTACGGN	530
QY	309	Arg 309	
Db	531	AAG 533	
RESULT 7			
BQ142519/c		2141 bp mRNA linear EST 24-APR-2002	
LOCUS		Contig6 Metarhizium anisopliae sf. acridum ARSEF 324 Metarhizium	
DEFINITION		anisopliae var. acridum cDNA, mRNA sequence.	
ACCESSION		BQ142519	
VERSION		1	
KEYWORDS		Metarhizium anisopliae var. acridum	
SOURCE		Metarhizium anisopliae var. acridum	
ORGANISM		Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Clavicipitaceae; Clavicipitaceae; Metarhizium	
REFERENCE		1 (bases 1 to 2141)	
AUTHORS		Frelmoser,F.M., Screen,S., Baga,S., Hu,G. and St. Leger,R.J.	
TITLE		EST analysis of genes expressed by two different insect pathogenic fungi during optimized secretion of proteins	
JOURNAL		Unpublished	
COMMENT		Contact: Frelmoser F. M. Department of Entomology University of Maryland 4112 Plant Sciences Building, College Park, MD 20742, USA Tel: 301 405 16 13 Fax: 301 314 92 90 Email: ff34@umail.umd.edu.	
FEATURES			
source			
Location/Qualifiers			
1. 2141			
/organism="Metarhizium anisopliae var. acridum"			
/mol_type="mRNA"			
/strain="ARSEF 324"			
/db_xref="taxon:92637"			
/clone_lib="Metarhizium anisopliae sf. acridum ARSEF 324"			
/note="Vector: Unizap; Metarhizium anisopliae sf. acridum was grown on insect cuticle and chitin for 24 hours. A cDNA library was constructed in the unidirectional Lambda vector Unizap."			

```

BASE COUNT      534 a   505 c   631 g   468 t       3 others
ORIGIN

Alignment Scores:
Pred. No.:      1,09e-10      Length:      2141
Score:          207.50      Matches:      73
Percent Similarity: 43.08%      Conservative: 36
Best Local Similarity: 28.85%      Mismatches: 91
Query Match:      9.23%      Indels:      53
DB:              13      Gaps:      12

US-09-985-689A-1 (1-434) x B0142519 (1-2141)
Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1059 GGTGAGGCTACTTGGGTATATATATATGACACTGGTGTGAGCTCTCC-----1012
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 1011 -----CACCCGAGCTTGGCGGTGCGCCCATCTGGCTCAGGAGCTTCATCAAGGTCAA 958
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 957 AAC---CGTGATGGCCAGCGCATCGGACTCACTCGCTGGTACTATT-----GGT 910
Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp--- 99
Db 909 AGCCGAAGCTACGGTGTGCCAAAATGCCAAGCTCTTGTCTGTCAGAGTCTTGTATGAC 850
Qy 100 ---SerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAla 118
Db 849 CAGGGCAGTGTTCTACTCGGATCATCATCGGATGCGAGCTTGTGTTGCCAGGACTCC 790
Qy 119 TyrSer-----AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsn 135
Db 789 AAGAGTCGTAACTGCCCAATGGCCACATTCCTCCATGAGTCTGGGA-----742
Qy 136 GlyAlaTyrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThr 155
Db 741 GGTGGCTACTCGGTCCTGCAACAGGGTGGCCCTGCTTGGTCAGGCTGGTGTCTC 682
Qy 156 IleLeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThr 175
Db 681 CTGTCGCTGCGCGCTGGCAGCATAAACCGGATGCCCAACACACTCT---CCCGCTTCC 625
Qy 176 AlaLysAsnAlaIleThrValGlyAlaThr-----GluAsnLeuArgProSerPheGly 193
Db 624 GAGCCTACTGCTGCACACTGTTGTGGCAGTCCGCTCAGATGACAGCCGATCTACCTTTCC 565
Qy 194 SerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAsp 213
Db 564 AACTAC-----559
Qy 214 GlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSer 233
Db 558 GGCAGAGTT---GTCGATATCTTCGCTCCTGTTACCGGATCTCTCCACC-----511
Qy 234 LeuAlaProAspSerSerPheThrAlaAsnHisAspSerLysTyrAlaTyrMetGlyGly 253
Db 510 -----TGGATCAAT-----GGCGGCACCAACACCATCTCTGGC 478
Qy 254 ThrSerMetAlaThrProIleValAlaGlyAsnValAla 266
Db 477 ACCTCCATGGCTACTCCCATATATGCTGGTCTTGCTGCC 439

RESULT 8
A0652212/c
LOCUS      508 bp      DNA      linear      GSS 22-JUN-1999
DEFINITION Sheared DNA-8F2.TF Sheared DNA Trypanosoma brucei genomic clone
ACCESSION A0652212
VERSION   A0652212.1 GI:5145398
KEYWORDS  GSS.

```

```

SOURCE      Trypanosoma brucei
ORGANISM    Trypanosoma brucei
REFERENCE   1 (bases 1 to 508)
AUTHORS    El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Sub,E., Malek,J., Fujii,C.,
            Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.,
            Fraser,C. and Adams,M.
TITLE       Determination of clone end sequences from Trypanosoma brucei GUTat
            10.1 sheared DNA library
JOURNAL     Unpublished
COMMENT     Other_GSSs: Sheared DNA-8F2.TR
            Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
            DNA library constructed at TIGR. Clones will be available for
            distribution through ATCC. Sheared DNA end sequences search page:
            http://www.tigr.org/tdb/mdb/tbdb/.
            Seq primer: M13-Forward
            Class: shotgun.
FEATURES    Location/Qualifiers
            1..508
            /organism="Trypanosoma brucei"
            /mol_type="genomic DNA"
            /strain="TREU927/4 GUTat 10.1"
            /db_xref="taxon:5691"
            /clone_lib="Sheared DNA-8F2"
            /notes="Vector: pUC18; Site 1: SmaI; Constructed at The
            Institute for Genomic Research (TIGR), Rockville, MD.
            Genomic DNA isolated from a cloned population of
            Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
            sheared to give a tight size distribution (approx 2 kb).
            The v + i method used for the library construction is
            described in detail in Smith, H.O. and Venter, J.C.
            (Making small insert libraries for whole genome shotgun
            sequencing projects. In Genome Sequencing: A Practical
            Approach, eds. M. Vaudin and B. Borell, Oxford University
            Press, 1999)."
BASE COUNT  130 a   135 c   129 g   114 t
ORIGIN

Alignment Scores:
Pred. No.:      7.73e-11      Length:      508
Score:          200.00      Matches:      57
Percent Similarity: 50.30%      Conservative: 28
Best Local Similarity: 33.73%      Mismatches: 56
Query Match:      8.90%      Indels:      28
DB:              28      Gaps:      7

US-09-985-689A-1 (1-434) x A0652212 (1-508)
Qy 156 IleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThr 175
Db 503 CTTATCTCTCCATGGCAACAGATATCCAAAGATGCC-----CTAATGACTCGCTGTCGT 450
Qy 176 AlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyr 195
Db 449 GGTAAGAAGCTGATGCTGGTGGGTCACACAAAAGCTG-----TTTGAGCTGTCG 399
Qy 196 AlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArg 215
Db 398 AAACAC-----ATTGTTTCTTCGTTTCTTCGATGTCCTCAACATACGACGCTAGG 348
Qy 216 IleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAla 235
Db 347 ATGAACCCGATCTTGTGCGTCCCGGGGAGAGGTTGGTGTGCTCTTCTCTGCAAA 288

```



```
Db 446 GAGCCTTCTGCTGCACTGTTGGTGCTGCGGAAATGACAGCGATCTTCCITCTTC 505
Qy 194 SerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAsp 213
Db 506 AACTAC----- 511
Qy 214 GlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSer 233
Db 512 GGCAGAGTT---GTCGATATTTTCGCTCTGCTGCAATGTCTTTCCACC----- 559
Qy 234 LeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGly 253
Db 560 -----TGGATTGGT-----GGCCGCACAAACACCATCTCTGGT 592
Qy 254 ThrSerMetAlaThrPro 259
Db 593 ACCTNCAATGGCTACTCCC 610
```

Search completed: July 28, 2003, 08:38:57
Job time : 2913.34 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 18:58:19 ; Search time 15.0173 Seconds
(without alignments)
1222.784 Million cell updates/sec

Title: US-09-985-689A-1

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQYNNVPGVQTFSLAIVN 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA.*

1: /cgn2.6/prodata/1/iaa/5A_COMB.pep.*

2: /cgn2.6/prodata/1/iaa/5B_COMB.pep.*

3: /cgn2.6/prodata/1/iaa/6A_COMB.pep.*

4: /cgn2.6/prodata/1/iaa/6B_COMB.pep.*

5: /cgn2.6/prodata/1/iaa/6C_COMB.pep.*

6: /cgn2.6/prodata/1/iaa/6D_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2247	100.0	640	4	US-09-509-814A-6
2	2242	99.8	640	4	US-09-509-814A-8
3	2183	97.2	639	4	US-09-509-814A-4
4	2155	95.9	639	4	US-09-509-814A-1
5	2155	95.9	640	4	US-09-509-814A-2
6	2125.5	94.6	641	2	US-08-873-479-42
7	1986.5	88.4	433	4	US-09-104-623A-4
8	1986.5	88.4	433	4	US-09-019-532-4
9	1986.5	88.4	635	2	US-08-873-479-43
10	1581.5	70.4	345	4	US-09-512-251A-10
11	1581.5	70.4	345	4	US-09-515-150A-10
12	452.5	20.1	659	3	US-08-894-818B-1
13	452.5	20.1	659	4	US-09-445-472-12
14	414	18.4	412	4	US-09-445-472-1
15	414	18.4	522	3	US-08-894-818B-3
16	414	18.4	522	4	US-09-445-472-4
17	414	18.4	654	3	US-08-894-818B-35
18	414	18.4	654	4	US-09-445-472-16
19	401	17.8	659	3	US-08-894-818B-5
20	346	15.4	520	3	US-09-000-016-7
21	346	15.4	520	4	US-09-514-340-7
22	346	15.4	734	3	US-09-000-016-4
23	346	15.4	734	4	US-09-514-340-4
24	346	15.4	823	3	US-09-000-016-2
25	346	15.4	823	4	US-09-514-340-2
26	304.5	13.6	903	1	US-08-750-532-1
27	304.5	13.6	1398	1	US-08-750-532-9

28	304.5	13.6	1398	3	US-08-894-818B-8	Sequence 8, Appli
29	304.5	13.6	1398	4	US-09-445-472-6	Sequence 6, Appli
30	295.5	13.2	237	1	US-08-750-532-18	Sequence 18, Appli
31	283.5	12.6	418	2	US-08-873-479-44	Sequence 44, Appli
32	276	12.3	418	4	US-09-966-921A-2	Sequence 2, Appli
33	256.5	11.4	397	4	US-08-328-352-7533	Sequence 7533, Ap
34	254.5	11.3	275	1	US-08-431-387-1	Sequence 1, Appli
35	251	11.2	280	1	US-08-434-255-8	Sequence 8, Appli
36	251	11.2	280	1	US-08-459-967-8	Sequence 8, Appli
37	251	11.2	280	1	US-08-460-327-8	Sequence 8, Appli
38	251	11.2	280	1	US-08-459-871-8	Sequence 8, Appli
39	251	11.2	280	3	US-09-024-532-2	Sequence 2, Appli
40	251	11.2	280	4	US-09-104-623A-2	Sequence 2, Appli
41	251	11.2	280	4	US-09-019-532-2	Sequence 2, Appli
42	251	11.2	280	4	US-09-417-359A-2	Sequence 2, Appli
43	251	11.2	370	1	US-08-434-255-6	Sequence 6, Appli
44	251	11.2	370	1	US-08-459-967-6	Sequence 6, Appli
45	251	11.2	370	1	US-08-460-327-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-509-814A-6
; Sequence 6, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUYOSHI
; APPLICANT: KOBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUMI
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 640
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-509-814A-6

Query Match	100.0%	Score 2247;	DB 4;	Length 640;
Best Local Similarity	100.0%;	Pred. No. 2e-173;		
Matches 434;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Y	1	NDVARGIVKADVAQSSYGLYGOGIIVAVADTGLDTRNDSSMHFAFRGKITALYALGRTN	60	
b	207	NDVARGIVKADVAQSSYGLYGOGIIVAVADTGLDTRNDSSMHFAFRGKITALYALGRTN	266	
Y	61	NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSAOYS	120	
b	267	NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSAOYS	326	
Y	121	AGARIHTNSWGAANGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI	180	
b	327	AGARIHTNSWGAANGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI	386	
Y	181	TVGATENLRPSFGSYADNINHVAFSSRGPTKGRKPKDVMAPCTFTLSARSLAPDSSF	240	
b	387	TVGATENLRPSFGSYADNINHVAFSSRGPTKGRKPKDVMAPCTFTLSARSLAPDSSF	446	
Y	241	WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLCY	300	

201405 SID 192
201405
201405

Tue Jul 29 14:12:41 2003

361 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGRNNVNFVINAPOSQTYTIEVQAYN 420
 567 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGRNNVNFVINAPOSQTYTIEVQAYN 626

447 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKAALIAAGADIGLY 506
 301 PNGNOGWRVTLDKSLNVAAYVNESSLSSTOKATYSFTATAGKPKLISLVMSDAPASTTA 360
 507 PNGNOGWRVTLDKSLNVAAYVNESSLSSTOKATYSFTATAGKPKLISLVMSDAPASTTA 566

361 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGRNNVNFVINAPOSQTYTIEVQAYN 420
 567 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGRNNVNFVINAPOSQTYTIEVQAYN 626

421 VPVGPOTFSLAIVN 434
 627 VPVGPOTFSLAIVN 640

RESULT 2
 US-09-509-814A-8

; Sequence 8, Application US/09509814A
 ; Patent No. 6376227
 ; GENERAL INFORMATION:
 ; APPLICANT: TAKAIWA, MIKIO
 ; APPLICANT: OKUDA, MITSUYOSHI
 ; APPLICANT: SAEKI, KATSUHIISA
 ; APPLICANT: KUBOTA, HIROMI
 ; APPLICANT: HITOMI, JUN
 ; APPLICANT: KAGEYAMA, YASUSHI
 ; APPLICANT: SHIKATA, SHITSUW
 ; APPLICANT: NOMURA, MASAFUMI
 ; TITLE OF INVENTION: ALKALINE PROTEASE
 ; FILE REFERENCE: 0327-0832-OPCT
 ; CURRENT FILING DATE: 2000-04-06
 ; CURRENT APPLICATION NUMBER: PCT/JP98/04528
 ; PRIOR FILING DATE: 1998-10-07
 ; PRIOR APPLICATION NUMBER: JP 9-274570
 ; PRIOR FILING DATE: 1997-06-08
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 8
 ; LENGTH: 640
 ; TYPE: PRT
 ; ORGANISM: Bacillus sp.
 US-09-509-814A-8

Query Match 99.8%; Score 2242; DB 4; Length 640;
 Best Local Similarity 99.8%; Pred. No. 4.9e-173;
 Matches 433; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
 Db 207 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 266
 Qy 61 NANTDNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
 Db 267 NANTDNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 326
 Qy 121 AGARHTNSGWAAGVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 Db 327 AGARHTNSGWAAGVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 386
 Qy 181 TVGATENLRPSFGSYADNINHVAFQSSRGPTKDGRIKPDVMAPGTFILSARSLAPDSSF 240
 Db 387 TVGATENLRPSFGSYADNINHVAFQSSRGPTKDGRIKPDVMAPGTFILSARSLAPDSSF 446
 Qy 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKAALIAAGADIGLY 300
 Db 447 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKAALIAAGADIGLY 506
 Qy 301 PNGNOGWRVTLDKSLNVAAYVNESSLSSTOKATYSFTATAGKPKLISLVMSDAPASTTA 360
 Db 507 PNGNOGWRVTLDKSLNVAAYVNESSLSSTOKATYSFTATAGKPKLISLVMSDAPASTTA 566

Qy 361 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGRNNVNFVINAPOSQTYTIEVQAYN 420
 Db 567 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGRNNVNFVINAPOSQTYTIEVQAYN 626
 Qy 421 VPVGPOTFSLAIVN 434
 Db 627 VPVGPOTFSLAIVN 640

RESULT 3
 US-09-509-814A-4

; Sequence 4, Application US/09509814A
 ; Patent No. 6376227
 ; GENERAL INFORMATION:
 ; APPLICANT: TAKAIWA, MIKIO
 ; APPLICANT: OKUDA, MITSUYOSHI
 ; APPLICANT: SAEKI, KATSUHIISA
 ; APPLICANT: KUBOTA, HIROMI
 ; APPLICANT: HITOMI, JUN
 ; APPLICANT: KAGEYAMA, YASUSHI
 ; APPLICANT: SHIKATA, SHITSUW
 ; APPLICANT: NOMURA, MASAFUMI
 ; TITLE OF INVENTION: ALKALINE PROTEASE
 ; FILE REFERENCE: 0327-0832-OPCT
 ; CURRENT APPLICATION NUMBER: US/09/509,814A
 ; CURRENT FILING DATE: 2000-04-06
 ; PRIOR APPLICATION NUMBER: PCT/JP98/04528
 ; PRIOR FILING DATE: 1998-10-07
 ; PRIOR APPLICATION NUMBER: JP 9-274570
 ; PRIOR FILING DATE: 1997-06-08
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 4
 ; LENGTH: 639
 ; TYPE: PRT
 ; ORGANISM: Bacillus sp.
 US-09-509-814A-4

Query Match 97.2%; Score 2183; DB 4; Length 639;
 Best Local Similarity 96.3%; Pred. No. 2.9e-168;
 Matches 418; Conservative 13; Mismatches 3; Indels 0; Gaps 0;

Qy 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
 Db 206 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 265
 Qy 61 NANTDNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
 Db 266 NANTDNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 325
 Qy 121 AGARHTNSGWAAGVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 Db 326 AGARHTNSGWAAGVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 385
 Qy 181 TVGATENLRPSFGSYADNINHVAFQSSRGPTKDGRIKPDVMAPGTFILSARSLAPDSSF 240
 Db 386 TVGATENLRPSFGSYADNINHVAFQSSRGPTKDGRIKPDVMAPGTFILSARSLAPDSSF 445
 Qy 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKAALIAAGADIGLY 300
 Db 446 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKAALIAAGADIGLY 505
 Qy 301 PNGNOGWRVTLDKSLNVAAYVNESSLSSTOKATYSFTATAGKPKLISLVMSDAPASTTA 360
 Db 506 PNGNOGWRVTLDKSLNVAAYVNESSLSSTOKATYSFTATAGKPKLISLVMSDAPASTTA 565
 Qy 361 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGRNNVNFVINAPOSQTYTIEVQAYN 420
 Db 566 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGRNNVNFVINAPOSQTYTIEVQAYN 625
 Qy 421 VPVGPOTFSLAIVN 434
 Db 626 VPVGPOTFSLAIVN 639

```
RESULT 4
US-09-509-814A-1
; Sequence 1, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIISA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUW
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-OPCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (23)..(23)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (29)..(29)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (32)..(32)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (46)..(46)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (47)..(47)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (53)..(53)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (70)..(70)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (74)..(74)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (89)..(89)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (102)..(102)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (105)..(105)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (128)..(128)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (130)..(130)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (131)..(131)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
```

```
; LOCATION: (132)..(132)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (133)..(133)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (146)..(146)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (148)..(148)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (160)..(160)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (165)..(165)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (172)..(172)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (183)..(183)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (187)..(187)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (188)..(188)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (189)..(189)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (194)..(194)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (286)..(286)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (306)..(306)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (324)..(324)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (369)..(369)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (431)..(431)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (501)..(501)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (531)..(531)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (541)..(541)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (584)..(584)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (591)..(591)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (592)..(592)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (594)..(594)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (595)..(595)
```

Tue Jul 29 14:12:41 2003

us-09-985-689a-1.ra

OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (596)..(596)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (611)..(611)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (632)..(632)
OTHER INFORMATION: Xaa is any amino acid
US-09-509-814A-1

Query Match 95.9%; Score 2155; DB 4; Length 639;
Best Local Similarity 96.3%; Pred. No. 5.3e-166;
Matches 418; Conservative 0; Mismatches 16; Indels 0; Caps 0;
QY 1 NDVARGIVKADVAQSSGLYGOGQIVAVADTGLDGRNDSSMHEAPRGKITALLYALGRTN 60
Db 206 NDVARGIVKADVAQSSGLYGOGQIVAVADTGLDGRNDSSMHEAPRGKITALLYALGRTN 265
QY 61 NANDTNGHGTAVAGSVLNGSTKMGMAPOANLVFQSIMDSXGGLGGLPSNLQTLFSQAYS 120
Db 266 NANDTNGHGTAVAGSVLNGSTKMGMAPOANLVFQSIMDSXGGLGGLPSNLQTLFSQAYS 325
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNGGTILSAPGTAKNAI 180
Db 326 AGARIHTNSGAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNGGTILSAPGTAKNAI 385
QY 181 TVGATENLRPSFGSVADNINHVAQFSRRGPTKGRIPKPDVMAFGTILSARSSLAPDSSF 240
Db 386 TVGATENLRPSFGSVADNINHVAQFSRRGPTKGRIPKPDVMAFGTILSARSSLAPDSSF 445
QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAAALAGADIGLY 300
Db 446 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAAALAGADIGLY 505
QY 301 PNGNCGWGRVTLKSLNVAYNVNESSLSSTOKATYFTATAGPKLISLVMSDAPASTTA 360
Db 506 PNGNCGWGRVTLKSLNVAYNVNESSLSSTOKATYFTATAGPKLISLVMSDAPASTTA 565
QY 361 SVTLVNDLVLITAPNGTOYVGNDFTSPIYNDNDGRNNVENVFNAPOSGTYYTIEVQAYN 420
Db 566 SVTLVNDLVLITAPNGTYVGNDFTSPIYNDNDGRNNVENVFNAPOSGTYYTIEVQAYN 625
QY 421 VPGVGPOTSLAIVN 434
Db 626 VPGVGPOTSLAIVN 639

RESULT 5
US-09-509-814A-2
Sequence 2, Application US/09509814A
Patent No. 6376227
GENERAL INFORMATION:
APPLICANT: TAKAIWA, MIKIO
APPLICANT: OKUDA, MITSUYOSHI
APPLICANT: SAEKI, KATSUHIISA
APPLICANT: KUBOTA, HIROMI
APPLICANT: HITOMI, JUN
APPLICANT: KAGEYAMA, YASUSHI
APPLICANT: SHIKATA, SHITSUW
APPLICANT: NOMURA, MASAFUMI
TITLE OF INVENTION: ALKALINE PROTEASE
FILE REFERENCE: 0327-0832-OPCT
CURRENT APPLICATION NUMBER: US/09/509,814A
CURRENT FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: PCT/JF98/04528
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: JP 9-274570
PRIOR FILING DATE: 1997-06-08
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2

LENGTH: 640
TYPE: PRT
ORGANISM: Bacillus sp.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (3)..(3)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (24)..(24)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (30)..(30)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (33)..(33)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (47)..(47)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (48)..(48)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (54)..(54)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (71)..(71)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (75)..(75)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (90)..(90)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (103)..(103)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (106)..(106)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (129)..(129)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (131)..(131)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (132)..(132)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (133)..(133)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (134)..(134)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (147)..(147)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (149)..(149)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (161)..(161)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (166)..(166)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (173)..(173)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (184)..(184)
OTHER INFORMATION: Xaa is any amino acid

NAME/KEY: misc_feature
LOCATION: (188)..(188)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (189)..(189)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (190)..(190)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (195)..(195)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (287)..(287)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (307)..(307)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (325)..(325)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (370)..(370)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (432)..(432)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (502)..(502)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (532)..(532)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (542)..(542)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (585)..(585)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (592)..(592)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (593)..(593)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (595)..(595)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (596)..(596)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (597)..(597)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (612)..(612)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (633)..(633)
OTHER INFORMATION: Xaa is any amino acid

US-09-509-814A-2

Query Match 95.9%; Score 2155; DB 4: Length 640;
Best Local Similarity 96.3%; Pred. No. 5.3e-166;
Matches 418; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 1 NDVARGIVADVAAQSSYGLYGQGIIVAVADTGLDGTGRNDSSMHEAFRGKITALYALGRTN 60
DB 207 NDVARGIVADVAAQSSYGLYGQGIIVAVADTGLDGTGRNDSSMHEAFRGKITALYALGRTN 266
QY 61 NNDTNGHGTAVAGSVLGNSTNKGAPQANLVFQSTMDSGGLGLPSNLQTLFSOAYS 120
DB 267 NNDTNGHGTAVAGSVLGNSTNKGAPQANLVFQSTMDSGGLGLPSNLQTLFSOAYS 326

QY 121 AGARIHTNSWGAANGAYTTDSRNVDDYVRKNDMTLLFAAGNECPNGGTTISAPGTAKNAI 180
DB 327 AGARIHTNSWGAANGAYTTDSRNVDDYVRKNDMTLLFAAGNECPNGGTTISAPGTAKNAI 386
QY 181 TVGATENLRPSFGSYADNTNHVAQFSSRGPTKDGRIKPDVMAFGTFTLSARSSLAPDSSF 240
DB 387 TVGATENLRPSFGSYADNTNHVAQFSSRGPTKDGRIKPDVMAFGTFTLSARSSLAPDSSF 446
QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALTAGAADTGLGY 300
DB 447 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALTAGAADTGLGY 506
QY 301 PNGNOGWGRVTLDKSLNVAYVNESSLSSTSQATYSFTATAGKPLKISLWSDAPASTTA 360
DB 507 PNGNOGWGRVTLDKSLNVAYVNESSLSSTSQATYSFTATAGKPLKISLWSDAPASTTA 566
QY 361 SVTLVNDLVLITAPNGTQYVGNDFSPYNDWNGRNNVNFVINAPOSQGTFTIEVOAYN 420
DB 567 SVTLVNDLVLITAPNGTQYVGNDFSPYNDWNGRNNVNFVINAPOSQGTFTIEVOAYN 626
QY 421 VPVGQTFSLAIVN 434
DB 627 VPVGQTFSLAIVN 640

RESULT 6

US-08-873-479-42
; Sequence 42, Application US/08873479
; Patent No. 5891701
; GENERAL INFORMATION:
; APPLICANT: Sloma, Alan
; APPLICANT: Lynne, Christanson
; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide
; TITLE OF INVENTION: Having Protease Activity
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5891701o No. 5891701disk of No. 5891701th America
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/873,479
; FILING DATE: 12-JUN-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Agis, Cheryl H
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 5251.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 641 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-873-479-42

Query Match 94.6%; Score 2125.5; DB 2: Length 641;
Best Local Similarity 93.5%; Pred. No. 1.3e-163;
Matches 406; Conservative 19; Mismatches 8; Indels 1; Gaps 1;

TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
STRAIN: Bacillus sp. Y
US-09-104-623A-4

Query Match 88.4%; Score 1986.5; DB 4; Length 433;
Best Local Similarity 87.3%; Pred. No. 1.3e-152;
Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
DB 1 NDVARGIVKADVAQNNFGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
QY 61 NANDNGHGTHTVAGSVLNGSTNKGMAQANLVFQSIMDSGGGLGGLPSNLQTLFQSAYS 120
DB 61 NANDNGHGTHTVAGSVLNGSTNKGMAQANLVFQSIMDSGGGLGGLPSNLQTLFQSAYS 120
QY 61 NASDPNGHGTHTVAGSVLNGSTNKGMAQANLVFQSIMDSGGGLGGLPSNLQTLFQSAYS 119
DB 61 NASDPNGHGTHTVAGSVLNGSTNKGMAQANLVFQSIMDSGGGLGGLPSNLQTLFQSAYS 119
QY 121 AGARIHTNSWGAAGVNGATYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
DB 121 AGARIHTNSWGAAGVNGATYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
QY 120 AGARIHTNSWGAAPVNGATYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 179
DB 120 AGARIHTNSWGAAPVNGATYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAQGTIFILSARSLAPDSF 240
DB 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAQGTIFILSARSLAPDSF 240
QY 180 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAQGTIFILSARSLAPDSF 239
DB 180 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAQGTIFILSARSLAPDSF 239
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
DB 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
QY 240 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 299
DB 240 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 299
QY 301 PNGNQGWRVTLTKSLNVAQFSSRGPTKDGRIKPDVMAQGTIFILSARSLAPDSF 360
DB 301 PNGNQGWRVTLTKSLNVAQFSSRGPTKDGRIKPDVMAQGTIFILSARSLAPDSF 360
QY 360 SYTLVNDLVLITAPNGQKYVGNDFSPYDNDGNNVNFVINAPOSQGTITIEVOAYN 420
DB 360 SYTLVNDLVLITAPNGQKYVGNDFSPYDNDGNNVNFVINAPOSQGTITIEVOAYN 420
QY 421 VPVGPQTFSLAIVN 434
DB 420 VPVGPQTFSLAIVH 433

RESULT 8
US-09-019-532-4
; Sequence 4, Application US/09019532B.
; Patent No. 6416756
; GENERAL INFORMATION:
; APPLICANT: Olsen, Arne Agerlin
; APPLICANT: Prent, Annette
; TITLE OF INVENTION: A Modified Enzyme for Skin Care
; FILE REFERENCE: 4922.204-US
; CURRENT APPLICATION NUMBER: US/09/019,532B
; CURRENT FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: 0038/97
; EARLIER FILING DATE: 1997-01-10
; EARLIER APPLICATION NUMBER: 0754/97
; EARLIER FILING DATE: 1997-06-25
; EARLIER APPLICATION NUMBER: 60/051,381
; EARLIER FILING DATE: 1997-07-07
; EARLIER APPLICATION NUMBER: PCT/DK98/000015
; EARLIER FILING DATE: 1998-01-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-019-532-4

Query Match 88.4%; Score 1986.5; DB 4; Length 433;
Best Local Similarity 87.3%; Pred. No. 1.3e-152;
Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
DB 209 NDVARGIVKADVAQNNFGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 268
QY 61 NANDNGHGTHTVAGSVLNGSTNKGMAQANLVFQSIMDSGGGLGGLPSNLQTLFQSAYS 120
DB 269 NANDNGHGTHTVAGSVLNGSTNKGMAQANLVFQSIMDSGGGLGGLPSNLQTLFQSAYS 327
QY 121 AGARIHTNSWGAAGVNGATYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
DB 328 AGARIHTNSWGAAPVNGATYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 387
QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAQGTIFILSARSLAPDSF 240
DB 388 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAQGTIFILSARSLAPDSF 447
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
DB 448 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 507
QY 301 PNGNQGWRVTLTKSLNVAQFSSRGPTKDGRIKPDVMAQGTIFILSARSLAPDSF 360
DB 508 PNGNQGWRVTLTKSLNVAQFSSRGPTKDGRIKPDVMAQGTIFILSARSLAPDSF 567
QY 361 SYTLVNDLVLITAPNGQKYVGNDFSPYDNDGNNVNFVINAPOSQGTITIEVOAYN 420
DB 568 SYTLVNDLVLITAPNGQKYVGNDFSPYDNDGNNVNFVINAPOSQGTITIEVOAYN 627
QY 421 VPVGPQTFSLAIVN 434
DB 628 VPVGPQTFSLAIVH 641

RESULT 7
US-09-104-623A-4
; Sequence 4, Application US/09104623A
; Patent No. 6303752
; GENERAL INFORMATION:
; APPLICANT: Olsen, Arne Agerlin
; APPLICANT: Fatum, Tine Muxoll
; APPLICANT: Deussen, Heinz-Josef
; APPLICANT: Roggen, Erwin Ludo
; TITLE OF INVENTION: A Modified Polypeptide
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 63037520 No. 6303752disk of No. 6303752th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/104,623A
; FILING DATE: 25-JUN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 5256.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; STRANDEDNESS:

```

QY 1 NDVARGIVKADVAQSSYGLYGQGOIVAVADTGLDTCRNDSSMHEAFRGKITALYALGRTN 60
Db 1 NDVARGIVKADVAQSSYGLYGQGOIVAVADTGLDTCRNDSSMHEAFRGKITALYALGRTN 60
QY 61 NNDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSSGGGLGGLPSNLQTLFQSOAYS 120
Db 61 NNDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSSGGGLGGLPSNLQTLFQSOAYS 120
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
Db 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
QY 181 TVGATENLRPSFGSIADPNHIAQFSSRGATROGRIKPDVTAPGTFFILSARSSLPDSSF 240
Db 181 TVGATENLRPSFGSIADPNHIAQFSSRGATROGRIKPDVTAPGTFFILSARSSLPDSSF 240
QY 241 WANHDSKYAYMGCTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
Db 241 WANHDSKYAYMGCTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
QY 301 PNGNOGWRVTLDKSLNVAQFSSRGATROGRIKPDVTAPGTFFILSARSSLPDSSF 360
Db 301 PNGNOGWRVTLDKSLNVAQFSSRGATROGRIKPDVTAPGTFFILSARSSLPDSSF 360
QY 361 SVTLVNDLVLITAPNGTQVGNDFTSYNDNNDGRNENNVFINAPQSGTYTIEVOAYN 420
Db 361 SVTLVNDLVLITAPNGTQVGNDFTSYNDNNDGRNENNVFINAPQSGTYTIEVOAYN 420
QY 421 VPVGPQTFSLAIVN 434
Db 421 VPVGPQTFSLAIVN 434

```

RESULT 9

```

US-08-873-479-43
; Sequence 43, Application US/08873479
; Patent No. 5891701
; GENERAL INFORMATION:
; APPLICANT: Sloma, Alan
; APPLICANT: Lyme, Christian
; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide
; TITLE OF INVENTION: Having Protease Activity
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 5891701o No. 5891701disk of No. 5891701th America
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/873,479
; FILING DATE: 12-JUN-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Agis, Cheryl H
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 5251.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 635 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

```

```

; TOPOLOGY: linear
; US-08-873-479-43

```

```

Query Match      88.4%; Score 1986.5; DB 2; Length 635;
Best Local Similarity 87.3%; Pred. No. 2.2e-152;
Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;

```

```

QY 1 NDVARGIVKADVAQSSYGLYGQGOIVAVADTGLDTCRNDSSMHEAFRGKITALYALGRTN 60
Db 203 NDVARGIVKADVAQSSYGLYGQGOIVAVADTGLDTCRNDSSMHEAFRGKITALYALGRTN 262
QY 61 NNDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSSGGGLGGLPSNLQTLFQSOAYS 120
Db 263 NNDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSSGGGLGGLPSNLQTLFQSOAYS 321
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
Db 322 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 381
QY 181 TVGATENLRPSFGSIADPNHIAQFSSRGATROGRIKPDVTAPGTFFILSARSSLPDSSF 240
Db 382 TVGATENLRPSFGSIADPNHIAQFSSRGATROGRIKPDVTAPGTFFILSARSSLPDSSF 441
QY 241 WANHDSKYAYMGCTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
Db 442 WANHDSKYAYMGCTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 501
QY 301 PNGNOGWRVTLDKSLNVAQFSSRGATROGRIKPDVTAPGTFFILSARSSLPDSSF 360
Db 502 PNGNOGWRVTLDKSLNVAQFSSRGATROGRIKPDVTAPGTFFILSARSSLPDSSF 561
QY 361 SVTLVNDLVLITAPNGTQVGNDFTSYNDNNDGRNENNVFINAPQSGTYTIEVOAYN 420
Db 562 SVTLVNDLVLITAPNGTQVGNDFTSYNDNNDGRNENNVFINAPQSGTYTIEVOAYN 621
QY 421 VPVGPQTFSLAIVN 434
Db 622 VPVGPQTFSLAIVN 635

```

RESULT 10

```

US-09-512-251A-10
; Sequence 10, Application US/09512251A
; Patent No. 6555355
; GENERAL INFORMATION:
; APPLICANT: Hansen, Peter
; APPLICANT: Bauditz, Peter
; APPLICANT: Mikkelsen, Frank
; APPLICANT: Andersen, Kim
; TITLE OF INVENTION: Protease Variants and Compositions
; FILE REFERENCE: 5349.204-US
; CURRENT APPLICATION NUMBER: US/09/512,251A
; CURRENT FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Bacillus
; US-09-512-251A-10

```

```

Query Match      70.4%; Score 1581.5; DB 4; Length 345;
Best Local Similarity 95.3%; Pred. No. 4.9e-120;
Matches 303; Conservative 10; Mismatches 4; Indels 1; Gaps 1;

```

```

QY 1 NDVARGIVKADVAQSSYGLYGQGOIVAVADTGLDTCRNDSSMHEAFRGKITALYALGRTN 60
Db 29 NDVARGIVKADVAQSSYGLYGQGOIVAVADTGLDTCRNDSSMHEAFRGKITALYALGRTN 88
QY 61 NNDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSSGGGLGGLPSNLQTLFQSOAYS 120
Db 89 NNDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSSGGGLGGLPSNLQTLFQSOAYS 147

```

QY 121 AGARIHNSWGAAYNGAYTTDSNRVDYVYKNDMTILFAAGNEGPNNGTTISAPGTAKNAI 180
DB 148 AGARIHNSWGAAYNGAYTTDSNRVDYVYKNDMTILFAAGNEGPNNGTTISAPGTAKNAI 207
QY 181 TVGATENLRPSGSGYADNHNHVAQFSSRGPTKGRKIPDYMAGPTFILSARSSSLAPDSSF 240
DB 208 TVGATENLRPSGSGYADNHNHVAQFSSRGPTKGRKIPDYMAGPTFILSARSSSLAPDSSF 267
QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAAIAGAADIGLGY 300
DB 268 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAAIAGAADVGLGF 327
QY 301 PNGNOGWGRVTLDKSLNV 318
DB 328 PNGNOGWGRVTLDKSLNV 345

RESULT 11
US-09-515-150A-10
; Sequence 10, Application US/09515150A
; Patent No. 6558938
; GENERAL INFORMATION:
; APPLICANT: Hansen, Peter
; APPLICANT: Bauditz, Peter
; APPLICANT: Mikkelsen, Frank
; APPLICANT: Andersen, Kim
; TITLE OF INVENTION: Protease Variants and Compositions
; FILE REFERENCE: 5348.204-US
; CURRENT APPLICATION NUMBER: US/09/515.150A
; CURRENT FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Bacillus
US-09-515-150A-10

Query Match 70.4%; Score 1581.5; DB 4; Length 345;
Best Local Similarity 95.3%; Pred. No. 4.9e-120; Indels 1; Gaps 1;
Matches 303; Conservative 10; Mismatches 4; Indels 1; Gaps 1;
QY 1 NDVARGIVKADVAGSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
DB 29 NDVARGIVKADVAGSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 88
QY 61 NANTNGHGHVAGSVLNGSTNKGMAQANLVFQSIMDSGGGLGSLPNSLOTLSOAYS 120
DB 89 NANTNGHGHVAGSVLNGSTNKGMAQANLVFQSIMDSGGGLGSLPNSLOTLSOAYS 147
QY 121 AGARIHNSWGAAYNGAYTTDSNRVDYVYKNDMTILFAAGNEGPNNGTTISAPGTAKNAI 180
DB 148 AGARIHNSWGAAYNGAYTTDSNRVDYVYKNDMTILFAAGNEGPNNGTTISAPGTAKNAI 207
QY 181 TVGATENLRPSGSGYADNHNHVAQFSSRGPTKGRKIPDYMAGPTFILSARSSSLAPDSSF 240
DB 208 TVGATENLRPSGSGYADNHNHVAQFSSRGPTKGRKIPDYMAGPTFILSARSSSLAPDSSF 267
QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAAIAGAADIGLGY 300
DB 268 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAAIAGAADVGLGF 327
QY 301 PNGNOGWGRVTLDKSLNV 318
DB 328 PNGNOGWGRVTLDKSLNV 345

RESULT 12
US-08-894-818B-1
; Sequence 1, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru

APPLICANT: MORISHITA, Mio
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MITA, Masanori
APPLICANT: ASADA, Kiyozo
APPLICANT: TSUNASAWA, Susumu
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894.818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 659 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-894-818B-1

Query Match 20.1%; Score 452.5; DB 3; Length 659;
Best Local Similarity 30.1%; Pred. No. 2e-28;
Matches 138; Conservative 67; Mismatches 153; Indels 101; Gaps 18;
QY 8 VKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALY-ALGRTNNANDTN 66
DB 145 IGADTVNNSLGVGSGVVAIVDTGIDAN-----HPDLKGVIGWDVAVNGRSTPYDDQ 198
QY 67 GHGTHVAGSVLNGSTNKGMAQANLVFQSIM--DSGGGLGSLPNSLOTLSOAYS 121
DB 199 GHGTHVAGSVLNGSTNKGMAQANLVFQSIM--DSGGGLGSLPNSLOTLSOAYS 258
QY 122 GARI-----HTNSWGAAYNGAYTTDSNRVDYVYKNDMTILFAAGNEGPNNGT 169
DB 259 GIRVINLSLGSQSSDGTSLSOAVNNANDA-----GIWCVAAAGSGNTYT 306
QY 170 ISAPGTAKNAITVAGTENLRPSGSGYADNHNHVAQFSSRGPTKGRKIPDYMAGPTFILS 229
DB 307 VGSPPAAASKVITVGA-----VDSNDNIASFSSRGPTADGRLKPEVAVPGVDIIA 355
QY 230 ARSSLAPDSSFWANHDSKYAYMGTSMTATPIVAG-NVAQLREHFVKNRGITPK--PSLLK 286
DB 356 PRAS---GTSMTGTPINDYTKASGTSMTATPIVAGSVGVGALILQAH-----PSWTPDKVK 404
QY 287 AALIAGA-----ADIGLGPNGNOGWGRVTLDKSL---NVAYVNESSSLSTTSOKATY 335
DB 405 TALLETADIVAPKEIADIAYGA-----GRVNYKAIKYDDYAKLITFTGSAVDKGSATH 457

Db 370 FEKGVYNPTDGTWIKVVSYS---GSANYQVDVVS 402
Search completed: July 25, 2003, 19:02:38
Job time : 17.0173 secs

CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 522 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
OTHER INFORMATION: /note= Xaa at position 428 is Gly or Val.
US-08-894-818B-3

Query Match 18.4%; Score 414; DB 3; Length 522;
Best Local Similarity 29.8%; Pred. No. 1.8e-25;
Matches 136; Conservative 59; Mismatches 147; Indels 114; Gaps 19;
QY 18 GLYGQGVAVADTGLDTRNDSMHEAFGRKITALYALGRNTNAN-----DTNGHGTH 71
Db 22 GYDGSGITGIIDTGID-----ASHPDLOGKV-----IGWVDFVNGRSYPYDDHGHGTH 70
QY 72 VAGSVLNGSTN---KGMAPQANLVQSIM--DSGGGLGLPSNLQTLFSQAYSAGARI 125
Db 71 VASIAAGTGAASNGKYKGMAPGAKLAGIKVLGADGSGSISTIIKGVWAVDNKOKYGIKV 130
QY 126 HTNSWGA-----AVNGAYTTDSRVNVDYVRKNDMTILFAAGNEGPNGGTISAP 173
Db 131 INLSGSSQSDGTDALSQAVNAAWDA-----GLVVVVAAGNSGPNKYTICSP 178
QY 174 GTAKNATVIGATENLRFSGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTFFLSARSS 233
Db 179 AAASKVITVGA-----VDKYDVITTSFSSRGPTADGRKPEVAVPAGNWIILAAARAS 227
QY 234 LAPDSSFWANHDSKYAYNGTSMATPIVAGNVAQLREHFVKNRGITPK--PSLLKAALIA 291
Db 228 -----GTSMGOPINDYITTAAPGTSMATPHVAGTAALLQ-----AHPSTPKVKITALIE 277
QY 292 GA-----ADIGLYPNQGWGRVTLDSLVNAYVNESSLSSTSQKA-----TYSFT 338
Db 278 TADIVKPEDEADIAYGA-----GRVNAYKAIN--YDNYAKLVFTGYVANKGSOHQFV 328
QY 339 ATAGKPLKISLVWSDAPASTASVTVNLDLVLITAPNGTQVGNDFTSFYNDNDWDCRNN 398
Db 329 ISGASFVTATLYWNNAN-----SDLDLYDPNGNQ--VDYSYTAIY-----G 369
QY 399 VENVFINAPOSQGTITIEVQAYNVPVGPQTFSIAVN 434


```

NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITALYALGRTNANDTNGHGHV
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITALYALGRTNANDTNGHGHV
X 10 20 30 40 50 60 70
80 90 100 110 120 130 140
AGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGPSNLQTLFSAQYAGARHNTNSWGAANGAYTTDSRN
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
AGSVLNGATNKGMAPOANLVFQSIMDSGGGLGPSNLQTLFSAQYAGARHNTNSWGAANGAYTTDSRN
80 90 100 110 120 130 140
150 160 170 180 190 200 210
VDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAITVGATENLRPSFGSYADNINHVAFSSRGPTKDGR
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
VDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAITVGATENLRPSFGSYADNINHVAFSSRGPTKDGR
150 160 170 180 190 200 210
220 230 240 250 260 270 280
KPDVMAPGTIFILSARSSSLAPDSSFWANHDSKYAYMGGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKAA
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
KPDVMAPGTIFILSARSSSLAPDSSFWANHDSKYAYMGGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKAA
220 230 240 250 260 270 280
290 300 310 320 330 340 350 360
LIAGAADIGLGPNGQGWGRVTLDKSLNVAYVNESSSLSTSOKATYSFTATAGKPLKISLVWSDAPASTTA
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
LIAGAADIGLGPNGQGWGRVTLDKSLNVAYVNESSSLSTSOKATYSFTATAGKPLKISLVWSDAPASTTA
290 300 310 320 330 340 350 360
SVTLVNDLVLITAPNGTQYVGNDFSPYNDNWDGRNNVENVFINAPQSGTYYTIEVOAYNVPVGPQTFSLAI
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
SVTLVNDLVLITAPNGTQYVGNDFSPYNDNWDGRNNVENVFINAPQSGTYYTIEVOAYNVPVGPQTFSLAI
370 380 390 400 410 420 430
SVTLVNDLVLITAPNGTQYVGNDFSPYNDNWDGRNNVENVFINAPQSGTYYTIEVOAYNVPVGPQTFSLAI
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
SVTLVNDLVLITAPNGTQYVGNDFSPYNDNWDGRNNVENVFINAPQSGTYYTIEVOAYNVPVGPQTFSLAI
370 380 390 400 410 420 430
X
VN
||
VN
X

```

3. US-09-985-689A-1 (1-434)

US-09-985-689A-6 Sequence 6, Application US/09985689A

Initial Score = 427 Optimized Score = 427 Significance = 1.00
 Residue Identity = 93% Matches = 406 Mismatches = 20
 Gaps = 0 Conservative Substitutions = 8

```

X 10 20 30 40 50 60 70
NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITALYALGRTNANDTNGHGHV
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITALYALGRTNANDTNGHGHV
X 10 20 30 40 50 60 70
80 90 100 110 120 130 140
AGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGPSNLQTLFSAQYAGARHNTNSWGAANGAYTTDSRN
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
AGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGPSNLQTLFSAQYAGARHNTNSWGAANGAYTTDSRN
80 90 100 110 120 130 140
150 160 170 180 190 200 210
VDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAITVGATENLRPSFGSYADNINHVAFSSRGPTKDGR
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
VDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAITVGATENLRPSFGSYADNINHVAFSSRGPTKDGR
150 160 170 180 190 200 210
220 230 240 250 260 270 280
KPDVMAPGTIFILSARSSSLAPDSSFWANHDSKYAYMGGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKAA
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
KPDVMAPGTIFILSARSSSLAPDSSFWANHDSKYAYMGGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKAA
220 230 240 250 260 270 280

```

```

290 300 310 320 330 340 350 360
LIAGAADIGLGPNGQGWGRVTLDKSLNVAYVNESSSLSTSOKATYSFTATAGKPLKISLVWSDAPASTTA
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
LIAGAADIGLGPNGQGWGRVTLDKSLNVAYVNESSSLSTSOKATYSFTATAGKPLKISLVWSDAPASTTA
290 300 310 320 330 340 350 360
SVTLVNDLVLITAPNGTQYVGNDFSPYNDNWDGRNNVENVFINAPQSGTYYTIEVOAYNVPVGPQTFSLAI
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
SVTLVNDLVLITAPNGTQYVGNDFSPYNDNWDGRNNVENVFINAPQSGTYYTIEVOAYNVPVGPQTFSLAI
370 380 390 400 410 420 430
SVTLVNDLVLITAPNGTQYVGNDFSPYNDNWDGRNNVENVFINAPQSGTYYTIEVOAYNVPVGPQTFSLAI
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
SVTLVNDLVLITAPNGTQYVGNDFSPYNDNWDGRNNVENVFINAPQSGTYYTIEVOAYNVPVGPQTFSLAI
370 380 390 400 410 420 430
X
VN
||
VN
X

```

4. US-09-985-689A-1 (1-434)

US-09-985-689A-7 Sequence 7, Application US/09985689A

Initial Score = 348 Optimized Score = 423 Significance = -0.61
 Residue Identity = 93% Matches = 406 Mismatches = 15
 Gaps = 1 Conservative Substitutions = 12

```

X 10 20 30 40 50 60 70
NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITALYALGRTNANDTNGHGHV
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITALYALGRTNANDTNGHGHV
X 10 20 30 40 50 60 70
80 90 100 110 120 130 140
AGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGPSNLQTLFSAQYAGARHNTNSWGAANGAYTTDSRN
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
AGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGPSNLQTLFSAQYAGARHNTNSWGAANGAYTTDSRN
80 90 100 110 120 130 140
150 160 170 180 190 200 210
VDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAITVGATENLRPSFGSYADNINHVAFSSRGPTKDGR
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
VDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAITVGATENLRPSFGSYADNINHVAFSSRGPTKDGR
150 160 170 180 190 200 210
220 230 240 250 260 270 280
KPDVMAPGTIFILSARSSSLAPDSSFWANHDSKYAYMGGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKAA
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
KPDVMAPGTIFILSARSSSLAPDSSFWANHDSKYAYMGGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKAA
220 230 240 250 260 270 280
290 300 310 320 330 340 350 360
LIAGAADIGLGPNGQGWGRVTLDKSLNVAYVNESSSLSTSOKATYSFTATAGKPLKISLVWSDAPASTTA
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
LIAGAADIGLGPNGQGWGRVTLDKSLNVAYVNESSSLSTSOKATYSFTATAGKPLKISLVWSDAPASTTA
290 300 310 320 330 340 350 360
SVTLVNDLVLITAPNGTQYVGNDFSPYNDNWDGRNNVENVFINAPQSGTYYTIEVOAYNVPVGPQTFSLAI
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
SVTLVNDLVLITAPNGTQYVGNDFSPYNDNWDGRNNVENVFINAPQSGTYYTIEVOAYNVPVGPQTFSLAI
370 380 390 400 410 420 430
SVTLVNDLVLITAPNGTQYVGNDFSPYNDNWDGRNNVENVFINAPQSGTYYTIEVOAYNVPVGPQTFSLAI
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
SVTLVNDLVLITAPNGTQYVGNDFSPYNDNWDGRNNVENVFINAPQSGTYYTIEVOAYNVPVGPQTFSLAI
370 380 390 400 410 420 430
X
VN
||
VN
VH
X

```